

**This Page Is Inserted by IFW Operations
and is not a part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- **BLACK BORDERS**
- **TEXT CUT OFF AT TOP, BOTTOM OR SIDES**
- **FADED TEXT**
- **ILLEGIBLE TEXT**
- **SKEWED/SLANTED IMAGES**
- **COLORED PHOTOS**
- **BLACK OR VERY BLACK AND WHITE DARK PHOTOS**
- **GRAY SCALE DOCUMENTS**

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

Ratio: 5.235 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

Alignment block:

US-09-673-254-1/rev x US-08-760-116-2 ..

Align seg 1/1 to: US-08-760-116-2 from: 1 to: 422

```

1351 GAGACGGGAGGAGCGCCCGGGTGGCCGTCGACCGCTTCGGTGCCTCAT 1302
      |||
1  ValSerGlyGluAlaProArgValAlaValAspProPheAlaLysProHe 17
1301 GATGACCATGACGCGCAAGCCCGAGGTGCACGACGCTTCCGGGAGCGCG 1252
      |||
17 tMetThrMetGlnArgLysProGluValHisAspAlaPheArgGluAlaG 34
51 GCGCGGTGCTCGAGGTGAACGCCCGCGCGGCGAGACCGCCCTGGGTGATC 1202
      |||
34 LProValValGluValAsnAlaProAlaGlyGlyProAlaTrpValIle 50
1201 ACCGATGACGCGCTCGCCCGCGAGGTGCTGGCCGATCCCGGGTGTGTA 1152
      |||
51 ThrAspAspAlaLeuAlaArgGluValLeuAlaAspProArgPheValIly 67
1151 GAGCCCGGACCTCGCCCGCGCTGGCGGGGGGTGAGCAGCGGTCTCG 1102
      |||
67 sAspProAspLeuAlaProAlaAlaTrpArgGlyValAspAspGlyLeuA 84
1101 ACATCCCGCTTCGGAGCTGGCTCCGTTTCACGCTCATGCGCCGTGAGCG 1052
      |||
84 sPLeProValProGluLeuArgProPheThrLeuIleAlaValAspGly 100
1051 GAGGGCCACGCGCGCTGGCGCGCATCCACGACCTGCTTCACCGCGCG 1002
      |||
101 GluAlaHisAlaArgLeuArgGlyIleHisAlaProAlaPheAsnProAr 117
1001 CCGGCTGGCGCGAGCGAGCGATCGCATCGCCGATCGCGCGCGCTGTC 952
      |||
117 gArgLeuAlaGluArgThrAspArgIleAlaAlaIleAlaGlyArgLeuL 134
951 TCACCGAACTCGCCGACGCGCTCGCGCGGTGGGCAACCGCGCGAGCTG 902
      |||
134 eutThrCluLeuAlaAspAlaSerGlyArgSerGlyLysProAlaGluLeu 150
901 ATCGGCGGCTTCGCTACCATTCCTCCGCTGTGTGTCATCTGGAGAGTGT 852
      |||
151 IleGlyGlyPheAlaTrpHisPheProLeuLeuValIleCysGluLeuLe 167
851 CGGTGTGCGCGTCCACCGATCGCGCGATGGCCCGGAGCGCGTCCAGCTTC 802
      |||
167 uGlyValProValThrAspProAlaMetAlaArgGluAlaValIleServa 184
801 TCAAGGCACTGGCGCTGGCGCGCGCGAGAGGGGGGGGTACGGGCAGC 752
      |||
184 euLysAlaLeuGlyLeuGlyGlyProGlnSerGlyGlyLysAspGlyThr 200
751 GACCTTGGCGGGGGGCGTGGCGACACTCGGCGCTGAGAGAGCTGCTCT 702
      |||
201 AspProAlaGlyGlyValProAspThrSerAlaLeuGlnSerLeuLeuLe 217
701 CGAAGCGGTGCACTAGCCCGCGGAGACACACACCCCGACCATGACCCGG 652
      |||
217 uGluAlaValHisSerAlaArgArgAsnAspThrProThrMetThrArgV 234
651 TGCCTGTAGAGCGCGCGAGCGCGAGTTGCTGGCTCTCCGACAGCAG 602
      |||
234 alLeuTrpGluArgAlaGlnAlaGlnPheGlySerValSerAspAspGln 250
601 CTCGTCTACATGATCAGCGGCTCATCTTCGCGGCGACAGACACACCGG 552
      |||
251 LeuValIlyMetIleThrCluLeuIlePheAlaGlyHisAspThrThrCl 267
551 CTCCTTCTCGGGCTTCGCTCGCGAGAGTCTCGCGCGCGCGCTCGCGG 502

```

```

|||||
267 ySerPheLeuGlyPheLeuLeuAlaGluValLeuAlaGlyArgLeuAlaA 284
501 CGGATGCGGACGACGAGACGCGCTCTCCGGTTCGTGTGAGAGCGCTGCGC 452
      |||
284 IaAspAlaAspGluAspAlaValAlaSerArgPheValGluGluAlaLeuArg 300
451 TACACCGCGCGGTGCGCTTACAGGTGTGTGAGGTTCGTGCGCACGAGGT 402
      |||
301 TyrHisProProValProIlyThrLeuTrpArgPheAlaIleThrGluVa 317
401 GACCATCGCGCGCGCTCGCGCTCGCCCGCGAGCGCGCGGTGCTGTGACA 352
      |||
317 LThrIleGlyGlyValArgLeuProArgGlyAlaProValLeuValAspI 334
351 TCGAGGCGACCAACACACCGCGCGCGCATCAGACGAGCGCGCGCGCTTC 302
      |||
334 IeGluGlyThrAsnThrAspGlyArgHisHisAspAlaProHisAlaPhe 350
301 CACCGGACCGTCCCTCGTGGCGCGGCTCACCTTCGCGGACGAGCGCGCA 252
      |||
351 HisProAspArgProSerTrpArgArgLeuThrPheGlyAspGlyProHi 367
251 CTACTGCATCGGGGAGACAGCTCGCCAGCTGAGTGCAGCAGATGATCG 202
      |||
367 sTyrCysIleGlyGluGlnLeuAlaGlnLeuGlnSerArgTrpMetIleG 384
201 GGTACTGCGCGACAGAGTTCCCGGAGCGCGGACTGGCGCTGCGCTRACG 152
      |||
384 LylValLeuArgSerArgPheProGluAlaArgLeuAlaValAlaProIlyAsp 400
151 GAGTTGCGGTGTGGCGGAGGCGCGGCGCGAGCGCGCGCTCACCGCAACT 102
      |||
401 GluLeuArgTrpCysArgGlySclValaGlnThrAlaArgLeuThrGluLe 417
101 GCCGCTGTGCTGCGC 86
417 uProValTrpLeuArg 422

seq_name: /cgn2_6/ptodate/1/iaa/5B_COMB.dep: US-09-096-982-9
seq_documentation_block:
: Sequence 9, Application US/09096982
: Patent No. 5962293
: GENERAL INFORMATION:
: APPLICANT: Strohl, William R.
: APPLICANT: Dickens, Michael L.
: TITLE OF INVENTION: METHOD OF PRODUCING DOXORUBICIN
: NUMBER OF SEQUENCES: 9
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: CALFE, HALTER & GRISWOLD
: STREET: 800 Superior Avenue, Suite 1400
: CITY: Cleveland
: STATE: Ohio
: COUNTRY: USA
: ZIP: 44114-2688
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/096,982
: FILING DATE:
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: Golrick, Mary E.
: REGISTRATION NUMBER: 34829
: REFERENCE/DOCKET NUMBER: 22727/00131
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 216-622-8458
: TELEFAX: 216-241-0816

```


NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: CALFEF, HALTER & GRISWOLD
STREET: 800 Superior Avenue, Suite 1400
CITY: Cleveland
STATE: Ohio
COUNTRY: USA
ZIP: 44114-2688
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/096,982
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Golrick, Mary E.
REGISTRATION NUMBER: 34829
REFERENCE/DOCKET NUMBER: 22727/00131
TELECOMMUNICATION INFORMATION:
TELEPHONE: 216-622-8458
TELEFAX: 216-241-0816
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 474 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-096-982-8

alignment_scores: Quality: 2093.50 Length: 432
 Ratio: 4.973 Gaps: 1
Percent Similarity: 97.454 Percent Identity: 93.056

alignment_block:
US-09-673-254-1/rev x US-09-096-982-8 ..

Align seg 1/1 to: US-09-096-982-8 from: 1 to: 474

```
1378 GTACCCCGCGCGGCGGCGG...TGCCCTGTGAGCGGCGAGCGCCCG 1332
||||| |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
43 ValProtyrGlyasnSerGluGlyCysLeuMetSerGlyGluAlaProAr 59
1331 GGTGGCCGTGACCGCTTCCGCGTCCATGATGACCATGCGACGCGAAG 1282
||||| |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
59 gValAlaValaAspProPheSerCysPrometMetThrMetGlnArgLysP 76
1281 CCGAGGTGACGACGCTTCCGCGAGGCGCGCGGCGTGTGAGGTGAAC 1232
||||| |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
76 roGluValaIHisAspAlaPheArgGluAlaGlyProValValGluValaAsn 92
1231 GCCCGCGGCGGCGACCGCGCTGTGATCATCACGATGACGCGCTCGCCG 1182
||||| |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
93 AlaProAlaGlyGlyProAlaTrpValIleThrAspAspAlaLeuAlaAr 109
1181 CGAGGTGCTGGCGGATCCCGGCTTCCGGAAGACCGCGACCTGCGCCCG 1132
||||| |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
109 gGluValaLeuAlaAspProArGlyValIysGlyProAspLeuAlaProT 126
1131 CCGCGTGGCGGGGGGTGAGCAGCGGTCTGACATCCCGCTTCCGAGACTG 1082
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
126 hrAlaTrpArgGlyValaAspAspGlyLeuAspIleProValProGluLeu 142
1081 CGTCCGTACGCTCATCGCGGTGAGCAGCGGAGCGCCACCGGCGCTGCG 1032
||||| |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
143 ArgProPheThrLeuIleAlaValaAspGlyGluAspHisArgArgLeuAr 159
1031 CCGCATTCACGACGCTGCTTCAACCGCGCGCGCTGGCGGAGCGAGCG 982
||||| |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
```

```
159 gArgIleHisAlaProAlaPheAsnProArgArgLeuAlaGluArgThra 176
981 ATCCGATCGCCGCGATGCGCGCGGCGTGCACACGAACTCGCGGAGCC 932
||||| |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
176 sPArgIleAlaAlaIleAlaAspArgLeuThrGluLeuAlaAspSer 192
931 TCGCGCGGTGCGGCAACCGCGCGAGCTGATCGCGGCTTCCGCTACCA 882
||| |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
193 SerAspArgSerGlyGluProAlaGluLeuIleGlyIlePheAlaTrpHi 209
881 CTTCGCCGCTGTGGTCATCTGCGAGCTGCTCGGTGTCGCGGTACCGATC 832
||||| |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
209 sPheProLeuLeuValIleCysGluLeuGlyValaProValThrAspR 226
831 CGCGGATGCGCGCGGAGCGGCTGACGCTTCTCAAGCGCACTCGCGCTGCG 782
||||| |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
226 roAlaMetAlaArgGluAlaValGlyValLeuYsaIleGlyLeuGly 242
781 GGCGCGAGAGCGGCGGGGTGAGCAGCGAGCGACCTCGGCGGCGTGGC 732
||||| |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
243 GlyProIleSerAlaGlyGlyAspGlyThrAspProAlaGlyAspValPr 259
731 GACACCTCGGCGCTGAGAGCGCTGCTCGAAGCGCGTCACTACGACCC 682
||||| |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
259 oAspThrSerAlaLeuGluIleuSerLeuLeuGluAlaValaIHisAlaAla 276
681 GGCGGAGACGACCGCCGACCATGACCGCGCGGTGTCAGAGCGCGCGAG 632
||||| |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
276 rArgGlyAspThrArgThrMetThrArgValLeuTrpGluArgAlaGln 292
631 GCCGAGTTGCGCTCGTCTCGAGCAGCACCTGCTTACATGATCAGCCG 582
||||| |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
293 AlaGluPheGlySerValSerAspAspGluLeuValIleThrMetIleThrGln 309
581 GCTCATCTTGGCGCGCCACGACACACCGCGCTCTTCTGAGGCTCTGCG 532
||||| |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
309 yLeuIlePheAlaGlyHisAspThrThrGlySerPheLeuGlyPheLeu 326
531 TCGGCGAGTCTCGCGCGCGCGCGCTCGCGGCGGATGCGCGAGGAGCGCC 482
||||| |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
326 euAlaGluValaLeuAlaGlyArgLeuAlaAlaAspAlaAspGlyAspAla 342
481 GTCTCCGCTTGTGTGAGGAGCGGCTGCTTACACCGCGCGGCTGCTTA 432
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
343 IleSerArgPheValGluGluAlaLeuArgHisIleProProValProTy 359
431 CACGTTGTGAGGCTGCTGCTGCGACGAGGTGACCATGCGCGGCTCGCGC 382
||||| |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
359 rSerLeuThrArgPheAlaAlaThrGluValaIleArgGlyValaArgL 376
381 TGCCTCCGCGAGCGCGCTGCTGTCGATCGAGCGGCGACACACCGAG 332
||||| |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
376 euProArgGlyAlaProValLeuValaAspIleGluGlyThrAspThrAsp 392
331 GCGCGCATCAGACGCGCGCGCAAGCGCTTCCACCGGAGCGCTCTGTG 282
||||| |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
393 GlyArgHisHisAspAlaProHisAlaPheHisProAspArgProSerArg 409
281 GCGCGGCTCACTTGGCGGAGCGCGCGCACTACTGATGCGGGAGAGAGC 232
||||| |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
409 gArgArgLeuThrPheGlyAspGlyProHisIleTyCysIleGlyGluIle 426
231 TCGCCGAGTGTGAGTGGCGAGCATGATGCGGCTGCTGCGTACGACAGTTG 182
||||| |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
426 euAlaGluLeuGluSerArgThrMetIleGlyValaLeuAlaGlySerArgPhe 442
181 CCGGAGCGCGACTGGCGCTGCGGTACGAGAGAGTGGGTGCGCGGAA 132
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
443 ProGluAlaArgLeuAlaValaProTyGluGluLeuArgTrpCysArgGly 459
131 GGGGCGCAAGCGCGGCGCTCAACGAACTGCGCGTCTGGCTGCGC 86
||||| |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
459 sGlyAlaGlnThrAlaArgLeuThrAspLeuProValIleTrpLeuArg 474
```



```

231  TCGCCACCTTGAGCGCGGCACAGTATGATCGGCGATACGCCACAGAGGTTG  162
426  euAlaInLeuGlnSerThrThreTlleGlyAlbLeuArgSerArgphe  442
181  CCCGAGCGCCACATGCGCGTCCCTACGACACAGATGGCGTGATCCCGCAA  132
443  ProGlnAlaArgLeuAlaValProTyrGlnIuLeuArgTyrPcySarGly  455
131  GGGGGGCCGAGACGGCGCGCTCACCGAACTGCCGTGGCTGGCGC  86
459  solYalagInThrAlaArgLeuThrAspLeuProValTyrPleuArg  474
seq_name: /cgm2_6/prodata/1/iaa/5B_COMB pep:US-09-096-962-5

```

```
|||||
334 1eGIuGIYThraSnThraSpGIYArgHISHisAspAlaProHISAlaPhe 350
301 CACCGGAGACGTCCTCGTGGGGGGGCTCACCTCGGGAGCGGGCGCA 252
351 HIsProAspArgProSerArgArgAlaGluThrPheGlyAspGIYProH 367
251 CTACTGCATCGGGAGAGACGTCGCCAGCTGGAGTCGCGCACGATGTCG 202
367 sTYrCysIleGIYGIuGIuInLeuAlaGIInLeuGIuSerArgThrMetIleG 384
201 GCGTACTGCGCAGACAGGTTCCCCGAGGGCCGACTGGCGCTGCCGTACGAC 152
384 lYValLeuArgSerArgPheProGInAlaArgLeuAlaValProTYrGIu 400
151 GAGTTCGGTGGTGGCGGAAGGGGGCCAGACGGCGCGGCTCACCGAACT 102
401 GIuLeuArgTYrPCysArgIYsGIYAlaGIInThrAlaArgLeuThraSpLe 417
101 GCCCGTCTGGTGGCGC 86
417 uProValTrPLeuArg 422
```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 11, 2002, 17:58:58 ; Search time 3477.18 Seconds
(Without alignments)
17272.376 Million cell updates/sec

Title: US-09-673-254-1
Perfect score: 2870
Sequence: 1 ggatccgacgggtacacg.....catgatgcgcgggtacacg/2870

Indexing table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Genbank
1: gb_ba:*
2: gb_hlg:*
3: gb_in:*
4: gb_ov:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_mu:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_hlg_hum:*
31: em_hlg_inv:*
32: em_hlg_other:*
33: em_hlg_inv:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result Query Match Length DB ID Description
SUMMARIES

1	2870	100.0	6985	1	SPU77891	077891 Streptomyces
2	2567.8	89.5	3196	1	SSU50973	U50973 Streptomyces
3	2542.2	88.6	3196	6	AR078146	AR078146 Sequence
4	2542.2	88.6	3196	6	AR083230	AR083230 Sequence
5	1732.4	80.4	1932	1	AF037708	AF037708 Sequence
6	1269	44.2	1269	6	AR078148	AR078148 Sequence
7	1221.6	42.6	2081	6	AR083232	AR083232 Sequence
8	1221.6	42.6	2081	6	AR078147	AR078147 Sequence
9	1102.2	38.4	3013	6	AR083231	AR083231 Sequence
10	1102.2	38.4	3013	6	AR075142	AR075142 Sequence
11	1046	36.4	1569	6	AR075142	AR075142 Sequence
12	1044.4	36.4	1569	6	AR144737	AR144737 Sequence
13	592	20.6	864	6	AR075141	AR075141 Sequence
14	592	20.6	864	6	AR144736	AR144736 Sequence
15	576.4	20.1	5985	1	SCU43704	SCU43704 Sequence
16	192.6	6.7	33320	1	SC9810	SC9810 Sequence
17	185	6.4	36028	1	SCB41	SCB41 Sequence
18	181.2	6.3	123580	1	AF263912	AF263912 Sequence
19	181.2	6.3	125401	6	AX211739	AX211739 Sequence
20	176	6.1	41807	1	SC6D11	SC6D11 Sequence
21	173.8	6.1	33810	1	SCB6	SCB6 Sequence
22	160.6	5.6	30941	1	SC4C6	SC4C6 Sequence
23	154.8	5.4	7870	1	STE18574	STE18574 Sequence
24	152.6	5.3	63734	1	AF127374	AF127374 Sequence
25	149.2	5.2	36583	1	SC5H1	SC5H1 Sequence
26	148.2	5.2	14750	1	SCF1	SCF1 Sequence
27	147.4	5.1	104326	1	AB070940	AB070940 Sequence
28	144.6	5.0	113193	1	AF357202	AF357202 Sequence
29	144.4	5.0	46166	1	SC8F11	SC8F11 Sequence
30	142	4.9	65140	6	AX211705	AX211705 Sequence
31	141.8	4.9	24268	1	AB070947	AB070947 Sequence
32	140.2	4.9	3912	1	RE0242746	RE0242746 Sequence
33	139.8	4.9	1257	6	AX097457	AX097457 Sequence
34	139.8	4.9	14759	1	AE004755	AE004755 Sequence
35	138.6	4.8	20535	1	SC2G4	SC2G4 Sequence
36	137.8	4.8	37750	1	SC6D7	SC6D7 Sequence
37	137.4	4.8	1470	1	AF087022	AF087022 Sequence
38	137.4	4.8	4257	6	AR144708	AR144708 Sequence
39	137.4	4.8	4257	6	AR1445616	AR1445616 Sequence
40	137.4	4.8	4342	1	AF079139	AF079139 Sequence
41	137.4	4.8	6633	14	HEHSV1G3	HEHSV1G3 Sequence
42	137.4	4.8	26245	14	HS1US	X06461 Herpes simp
43	137.4	4.8	26245	14	HS1US	L00036 Human herpe
44	137.4	4.8	152261	14	HE1CG	X14112 Herpes simp
45	137.4	4.8	152261	14	HE1CG	X14112 Herpes simp

ALIGNMENTS

RESULT 1
SPU77891
LOCUS 6985 bp DNA linear BCT 13-JAN-1999
DEFINITION Streptomyces peuceletius daunorubicin C-14 hydroxylase (doha),
doxorubicin biosynthesis enzyme DnrV (dnrv), daunorubicin C-13
ketoreductase (dnrv), daunorubicin acyl carrier protein (dpsg),
daunorubicin biosynthesis enzyme (dpsg), doxorubicin biosynthesis
protein Dnmr (dnmr), and putative daunorubicin biosynthesis protein
(dnrv) genes, complete cds.
U77891
U77891.1 GI:3778994

ACCESSION U77891

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL MEDLINE
U77891.1 178 (24), 7316-7321 (1996)
97113559

Streptomyces peuceletius
Streptomyces peuceletius
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
1 (bases 3983 to 6985)
Scott, C. and Hutchinson, C.R.
Enhanced antibiotic production by manipulation of the Streptomyces
peuceletius dnmr and dnmr genes involved in doxorubicin (adriamycin)
biosynthesis

REFERENCE 2 (bases 1 to 6985)
 Lomovskaya, N., Otten, S. L., Doi-Katayama, Y., Fonstein, L., Liu, X.-C., Takatsu, T., Inventi, A., Filippi, S., Torti, F., Colombo, A. L. and Hutchinson, C. R.
 Doxorubicin overproduction in Streptomyces peuceletii: cloning and characterization of the *dnrv* ketoreductase and *dnrv* genes and the *doxa* cytochrome P-450 hydroxylase gene
 J. Bacteriol. 181 (1), 305-318 (1999)

JOURNAL MEDLINE
 3 (bases 3893 to 6985)
 Scotti, C. and Hutchinson, C. R.
 Direct Submission
 Submitted (12-NOV-1996) School of Pharmacy, University of Wisconsin, 425 N. Charter St., Madison, WI 53706, USA
 4 (bases 1 to 6985)
 Lomovskaya, N., Otten, S. L., Doi-Katayama, Y., Fonstein, L., Liu, X.-C., Takatsu, T., Inventi, A., Filippi, S., Torti, F., Colombo, A. L. and Hutchinson, C. R.
 Direct Submission
 Submitted (21-OCT-1998) School of Pharmacy, University of Wisconsin, 425 N. Charter St., Madison, WI 53706, USA
 Sequence update by submitter
 On Oct 22, 1998 this sequence version replaced gi:1684919.

MARK LOCATION/QUALIFIERS
 1. 6985

gene
 CDS
 /organism="Streptomyces peuceletii"
 /strain="ATCC 29050"
 /db_xref="ATCC:29050"
 /db_xref="taxon:1950"
 complement(83..1330)
 /gene="doxa"
 complement(83..1330)
 /gene="CYP450 hydroxylase"
 /note="CYP450 hydroxylase"
 /codon_start=1
 /transl_table=1
 /product="daunorubicin C-14 hydroxylase"
 /protein_id="AAD04715.1"
 /db_xref="GI:3778995"
 /translation="MAVDPFACPMPTMQRREVDHAFREAGPVVEVNAAPGPAWITDALARVLADPREVDKDPDLPAAPRGVDDLDIPPELRFETLIAVDGEHRLRITHAPFNRLAEERDRIATAGRLTELADASGRSGKPAELIGFAAHFPLVYICELGVPYDPAARAVSVLKAIGGSGSGDGPAGVDTSLLESLLEBNASARKNDPMTRVLYLKERAGAEFSGVSDQLYMTTGLTFAGHDITGSLGLLEVLAVGRIADADEAVSRVEALRHPPEVPTLMRFATEVTTIGVRLPGAPVLDIEGNTDGRHDAFPAHPRDRSMRRLTFEGDGHICGLOLAQLSRTMIGVLRSEFPEARLAVYDELRLCKGAOTATLTELPMWR"
 complement(1348..2175)
 /gene="dnrv"
 complement(1348..2175)
 /gene="dnrv"
 /codon_start=1
 /transl_table=1
 /product="doxorubicin biosynthesis enzyme Dnrv"
 /protein_id="AAD04716.1"
 /db_xref="GI:3778996"
 /translation="MTRFAPGAPMPDLGSPVDAASADFTYGLFGWTATVSDPGAGYTFSSDGKLVAAVARHQIDTPYHRPFGNDQGMFAITVYFATDDADALTRKETAGEVINTPMDVYLGIRMAVEADPAGAPAFVWRKGVMEAGVTVGSGVSMVELVTDGIGARDFYPATLGLADPATGLKGVTPVWHIGDTPVAGTQELGVTGAVRHHMVLAVYHDCDAVRRARVLEGGSSANERADTPRRRRADLLDPHGAGTSVVELREGVYAAAGCAS"
 complement(2279..3142)
 /gene="dnrv"
 complement(2279..3142)
 /gene="dnrv"
 /codon_start=1
 /transl_table=1
 /product="daunorubicin C-13 ketoreductase"
 /protein_id="AAD04717.1"
 /db_xref="GI:3778997"
 /translation="MTASTPHAGTTPRGGLSGRTVLVTGATSGIGRAALAVAROGARVVLAVGDPERLRTVNEVARTAGAPADPAFELROVRLGRLDRYPRIDVMSNAGMFWSTRITTDGFEATIGVNLHAGFLARLRERLAGRLILTSDDATYQGRIDPD

gene
 CDS
 /organism="Streptomyces peuceletii"
 /strain="ATCC 29050"
 /db_xref="ATCC:29050"
 /db_xref="taxon:1950"
 complement(83..1330)
 /gene="doxa"
 complement(83..1330)
 /gene="CYP450 hydroxylase"
 /note="CYP450 hydroxylase"
 /codon_start=1
 /transl_table=1
 /product="daunorubicin C-14 hydroxylase"
 /protein_id="AAD04715.1"
 /db_xref="GI:3778995"
 /translation="MAVDPFACPMPTMQRREVDHAFREAGPVVEVNAAPGPAWITDALARVLADPREVDKDPDLPAAPRGVDDLDIPPELRFETLIAVDGEHRLRITHAPFNRLAEERDRIATAGRLTELADASGRSGKPAELIGFAAHFPLVYICELGVPYDPAARAVSVLKAIGGSGSGDGPAGVDTSLLESLLEBNASARKNDPMTRVLYLKERAGAEFSGVSDQLYMTTGLTFAGHDITGSLGLLEVLAVGRIADADEAVSRVEALRHPPEVPTLMRFATEVTTIGVRLPGAPVLDIEGNTDGRHDAFPAHPRDRSMRRLTFEGDGHICGLOLAQLSRTMIGVLRSEFPEARLAVYDELRLCKGAOTATLTELPMWR"
 complement(1348..2175)
 /gene="dnrv"
 complement(1348..2175)
 /gene="dnrv"
 /codon_start=1
 /transl_table=1
 /product="doxorubicin biosynthesis enzyme Dnrv"
 /protein_id="AAD04716.1"
 /db_xref="GI:3778996"
 /translation="MTRFAPGAPMPDLGSPVDAASADFTYGLFGWTATVSDPGAGYTFSSDGKLVAAVARHQIDTPYHRPFGNDQGMFAITVYFATDDADALTRKETAGEVINTPMDVYLGIRMAVEADPAGAPAFVWRKGVMEAGVTVGSGVSMVELVTDGIGARDFYPATLGLADPATGLKGVTPVWHIGDTPVAGTQELGVTGAVRHHMVLAVYHDCDAVRRARVLEGGSSANERADTPRRRRADLLDPHGAGTSVVELREGVYAAAGCAS"
 complement(2279..3142)
 /gene="dnrv"
 complement(2279..3142)
 /gene="dnrv"
 /codon_start=1
 /transl_table=1
 /product="daunorubicin C-13 ketoreductase"
 /protein_id="AAD04717.1"
 /db_xref="GI:3778997"
 /translation="MTASTPHAGTTPRGGLSGRTVLVTGATSGIGRAALAVAROGARVVLAVGDPERLRTVNEVARTAGAPADPAFELROVRLGRLDRYPRIDVMSNAGMFWSTRITTDGFEATIGVNLHAGFLARLRERLAGRLILTSDDATYQGRIDPD

gene
 CDS
 /organism="Streptomyces peuceletii"
 /strain="ATCC 29050"
 /db_xref="ATCC:29050"
 /db_xref="taxon:1950"
 complement(83..1330)
 /gene="doxa"
 complement(83..1330)
 /gene="CYP450 hydroxylase"
 /note="CYP450 hydroxylase"
 /codon_start=1
 /transl_table=1
 /product="daunorubicin C-14 hydroxylase"
 /protein_id="AAD04715.1"
 /db_xref="GI:3778995"
 /translation="MAVDPFACPMPTMQRREVDHAFREAGPVVEVNAAPGPAWITDALARVLADPREVDKDPDLPAAPRGVDDLDIPPELRFETLIAVDGEHRLRITHAPFNRLAEERDRIATAGRLTELADASGRSGKPAELIGFAAHFPLVYICELGVPYDPAARAVSVLKAIGGSGSGDGPAGVDTSLLESLLEBNASARKNDPMTRVLYLKERAGAEFSGVSDQLYMTTGLTFAGHDITGSLGLLEVLAVGRIADADEAVSRVEALRHPPEVPTLMRFATEVTTIGVRLPGAPVLDIEGNTDGRHDAFPAHPRDRSMRRLTFEGDGHICGLOLAQLSRTMIGVLRSEFPEARLAVYDELRLCKGAOTATLTELPMWR"
 complement(1348..2175)
 /gene="dnrv"
 complement(1348..2175)
 /gene="dnrv"
 /codon_start=1
 /transl_table=1
 /product="doxorubicin biosynthesis enzyme Dnrv"
 /protein_id="AAD04716.1"
 /db_xref="GI:3778996"
 /translation="MTRFAPGAPMPDLGSPVDAASADFTYGLFGWTATVSDPGAGYTFSSDGKLVAAVARHQIDTPYHRPFGNDQGMFAITVYFATDDADALTRKETAGEVINTPMDVYLGIRMAVEADPAGAPAFVWRKGVMEAGVTVGSGVSMVELVTDGIGARDFYPATLGLADPATGLKGVTPVWHIGDTPVAGTQELGVTGAVRHHMVLAVYHDCDAVRRARVLEGGSSANERADTPRRRRADLLDPHGAGTSVVELREGVYAAAGCAS"
 complement(2279..3142)
 /gene="dnrv"
 complement(2279..3142)
 /gene="dnrv"
 /codon_start=1
 /transl_table=1
 /product="daunorubicin C-13 ketoreductase"
 /protein_id="AAD04717.1"
 /db_xref="GI:3778997"
 /translation="MTASTPHAGTTPRGGLSGRTVLVTGATSGIGRAALAVAROGARVVLAVGDPERLRTVNEVARTAGAPADPAFELROVRLGRLDRYPRIDVMSNAGMFWSTRITTDGFEATIGVNLHAGFLARLRERLAGRLILTSDDATYQGRIDPD

BASE COUNT
 ORIGIN
 958 a
 2535 c 2543 g 949 t

DLNDRHRYSAAGAYCTSKOANIMTAARARMPDVAVSHRPGVRRIGRGVASS
 YFRNPFIRSAKAGADTVLWLASAPAEELTTGGTSDRRLSPVSGPTADAGLAALWE
 AGAAVAGDTAH
 complement(3139..3393)
 /gene="qpsc"
 complement(3139..3393)
 /gene="qpsc"
 /note="component of daunorubicin/doxorubicin polyketide
 synthase"
 /codon_start=1
 /transl_table=1
 /product="daunorubicin acyl carrier protein"
 /protein_id="AAD04718.1"
 /db_xref="GI:3778998"
 /translation="MAELSTELREIMKOSIGEDVDPDLADDTVTEEDGLDSLAVL
 ETVNHEKTYGVKLPHEELAVRPHSMILFVNRRLRAA"
 3493..4077
 /gene="qpsH"
 3493..4077
 /gene="qpsH"
 /codon_start=1
 /transl_table=1
 /product="daunorubicin biosynthesis enzyme"
 /protein_id="AAD04719.1"
 /db_xref="GI:3778999"
 /translation="MDRRLSALPDQARNRCTAPAGRRTRPGDRAARHHTTQGYTM
 DALPADAPTESADCTDPGTYLDVQVFRAROMLDDGRREDARTFADQGEVVG
 EAYGAGADTARARRTDRPAEGITRRHMGTLVYSGKPDDEVARSAYAVVLETPRDG
 DVLRLSTVCDDILVREAGWVRNHRVTRDGLD"
 4129..5646
 /gene="dnmt"
 4129..5646
 /gene="dnmt"
 /note="catalyzes unknown step in biosynthesis of
 daunosamine, an intermediate of doxorubicin biosynthesis;
 deoxysugar biosynthesis enzyme"
 /codon_start=1
 /transl_table=1
 /product="doxorubicin biosynthesis protein Dnmt"
 /protein_id="AAD04723.1"
 /db_xref="GI:1684920"
 /translation="MTAOLARSVLARDGLSGNDREWMAADRSAAQVNRHTERPLDD
 LKMSRHPTGVTSHTHGREFSIEGLDVHLPAGAPVPMWQPIVQAPGIEGLVYKER
 HGVLHCLVQAKFEGNPGGLQSLPTVQATSNTRVYGGKAIPLYLEHFRPAEROVVA
 DIVLOSEGSEFNRKRNMTVOYIDEVPLHEDFRMLTLGOLHLLGVENTYNNDAITY
 LACLPASGAEHPLADAPACADGPGGHRVVPARPDGCGRVDTCGRHSLVNSCAA
 EGSLSHYDVLVSMTADLRSTVDVTRPALNLAHPMTERDGAIAHESGRLEVMADV
 TPASREVPWSQPIEDKDGVAFLVRRIDGVLVHVAHARVEGVVDVIEADPTVOC
 TPLSLHALPAEARPRFLDAVLEAPPEVRVATVLAEGGRFYAVNTVMIVADHDIP
 DGEYRWLTLLHQVLGLRHSYVNVQARTLVACLSLSVSGSPVTRSAVPAPRRR"
 5654..6985
 /gene="dnrH"
 5654..6985
 /gene="dnrH"
 /note="catalyzes formation of baumycins, C-4' glycosides
 of doxorubicin; putative glycosyl transferase; DnrH"
 /codon_start=1
 /transl_table=1
 /product="putative baumycin biosynthesis protein"
 /protein_id="AAD04724.1"
 /db_xref="GI:1684921"
 /translation="MRVLFAATMAARSHVYAQVTLASALRTAGHEVLVASOPDVLDIV
 RAGLTAVRIGEDLNIEETREANASEEDRNGLGAMSNSRDPFPMDHLAGMTMT
 AMVFNWCEPWPVDLVGLARDWRPDLVLAGPVAARISGAARILPROM
 GMRRTAFRLILDRPPELBDPLAEMLTWTLEKCGSAGDSEVLGAGTIDTPPS
 MRPLDLPVPAVATVPYNGPSLGLWLRPRPRRLCTLGVSLGAEATGAGTAAVD
 VLAAGDLVEVAVATLRELRELGTLPANVRAVDVPPNALPSCSGIIHGGSGCTF
 MTAIVHAPOLIVPDMMDAMEKAHGLASGGSYVDARVDSPELLREVERIALDDPS
 YAAGARRVAEITGTSPNDIVYVLERLTAHEHRAGACGGPALKSPSTGA"

[illegible]

JOURNAL Submitted (08-MAR-1996) William R. Strohl, Ohio State University,
484 West 12th Avenue, Columbus, OH 43210 USA

FEATURES
source

1. .3196
/organism="Streptomyces sp."
/strain="C5"
/db_xref="taxon:1931"
33. .350
/function="unknown ketoreductase"
/note="orf1"
/codon_start=1
/transl_table=1
/protein_id="AAB08047.1"
/db_xref="GI:1354267"
/translation="YPRIDVAGNAGCMFMSRTTQDGEATLQVNHLAGFLTLRLHL
ERLAGEGLITLSDAAYTGRIQDPLDNDHRRHSAGCAVGTSGKANTATPEAARRMF
DLVLYSVHGEVFRIRIGCYASTYFRNPFLRSNAKADTLVLLAAAPAEELITGGT
YSDRLSPVSGPFTADGLIAKLMEASAAYADTAA"
663. .1501
/gene="orfa"
/note="orf1"
663. .666
/gene="orfa"
674. .1501
/gene="orfa"
674. .1501
/gene="orfa"
/function="unknown"
/codon_start=1
/transl_table=1
/protein_id="AAB08048.1"
/db_xref="GI:1354268"
/translation="MTFRAPGAPAMFDLGSPPVAASADFTYGLFGMTATVYSDGAGG
YTTSSDGKTLAAVAARHIDTPYHRRPGPDQGMPIATVYTRINDADALTRVET
AGGVYIMPMVDLGLGRNAVFPADSGAFAVWRKGVMEGAEVTPSSGVNVELTDD
IGVLRGVRATVGLIAPADTGRKGVTDVPMHLDHPPVAGTRGLTGAARPMAAVLSY
HDCATVRAAEALCGSVENEDVTPRGRRADLDPHGAGSFVELRELYPAAADGAS
14398. .2766
/gene="doxa"
1438. .2766
/gene="doxa"
/function="hydroxylation of C-14 of daunomycin to produce
doxorubicin"
/codon_start=1
/transl_table=1
/evidence="experimental"
/product="daunomycin C-14 hydroxylase"
/protein_id="AAB08049.1"
/db_xref="GI:1354269"
/translation="MSEGAPRVAADVPCSPMMTQRRPEVHADPREAGPVVEVNA
GPAVITIDDALAREVLADPRFKDPLAPAMRGVDGIDIPBELRPTLIAVGEDE
HRRARITHAPAFNRRLAERTDIRIAADRLTLADSDSGEAPLEIGFEVHNPFI
LVICELGVPVTPDPMAREAVGLKALIGGPOSAGSDGDPADGVDPTLSLELLE
AVHARRKDPRTMTRVLYXERQAEGFSYSDQLVYMTTGLIFASHDPTGSLGLVLA
VLARLADADGDAISRVEBALRHHPPVPTTLWRFAATEVITNGVRLPRCAVLDVI
EGVTIDRHHDAHPAHPDRPSRRRLTFGDDPHYCIGEQLAQLSRTMIGVLSRSPQ
ARLVAPEELRMCCKGAOTARLTDLPLWLR"
3038. .3041
/gene="daut"
3038. .3196
/gene="daut"
3050. .>3196
/gene="daut"
/function="putative transcriptional activator"
/codon_start=1
/transl_table=1
/product="daut"
/protein_id="AAB08050.1"
/db_xref="GI:1354270"
/translation="MQINMLGRLVAHHNGTSVPIAKKPROVESLALQAGTVVPVPA
LMEEL"

BASE COUNT 476 a 1184 c 1079 g 457 t

ORIGIN

Query Match 89.5%; Score 2567.8; DB 1; Length 3196;

Best Local Similarity 94.1%; Pred. No. 9,5e-205;
Matches 2697; Conservative 0; Mismatches 147; Indels 21; Gaps 2;

[illegible]

D	1825	GGATGCGCGGACAGACGCGCGGTGCTCTCAACGCTCCACGGCGATGAGCGTGAACGGACGCA	1766
Q	1084	gctccggaacggggagatgltcgaagaccgtctgtccaccgcccgccgaagcgggcggaagt	1143
D	1765	GCCTCGGAACGGGGATGTGAGACCGTGTCCACCCCCCGCAGCGGGTGGCGGAGAT	1706
Q	1144	cggggtctctcaagaaecggggatcggccagcacctcgcggcggaaggcgtlcatcgttga	1203
D	1705	CGGGGTCTTTCACGAACCGGGGATGGGCGACGACCTCGGGGCGAGGGCGTTCATGGGTGA	1646
Q	1204	tcaccgaaggcgggttcggcccgcgggggggtgttaacttcgagagacggggccgctccgga	1263
D	1645	TCACCAAGCGGGGTCCGGCCGGGGGGGCTTACCTTCAGACGAGGGGCCGCTTCCGA	1586
Q	1264	aggcgtctcgtgacacctcgggtctgctgtcatgtgltcatactgagaaacgcaagcgtgtcga	1323
D	1585	ATGCGTGTGCACTCGGGTTTGGCGTGTATGGTTCATCATGGGACACGAAAGGGGTGGA	1526
Q	1324	cgggcaacccggggcgctctgcgcgtcaagaggaacccgcgcgcgggggttacccctc	1383
D	1525	CGGCAACCGCGCGCGCTTCGCCGTCATGAGCACCGTCCGCGCGGGGTAGCGCTC	1466
Q	1384	ccggaagtgcgaacaacggaagaacggcccgccggtggtgtcgaagaagtccgcgcgcgc	1443
D	1465	CCGCAATTGACACCGAAGAACCGGGCCCCGTTGCGGGTGGACAGAGTCCGCCCCGCGCC	1406
Q	1444	ccctggcgtgtcggcggtctcgtctcgcagagacggcgcaagtctcaacggcgcgccgag	1503
D	1405	CGTGGGGGTGTGAGAGGGGTGCTTCTTCAGAGGAGCGCGCGAATTGAGCGCGCGCGGAC	1346
Q	1504	cgctcgctgcgaagtcgttgacggcggaacgaacggcccaagtgcggtacgcgcgcgt	1563
D	1345	CGTGGGTGTGCAGTGCTGTCAGCAGAAACAGCACGCGCCAGTGGGGCGGTACCGCGCGGT	1286
Q	1564	ggcgcgcgaactctgtgtgtgtgcgggaacgggtgtgttcaacgaatgtccgaacgggttcgt	1623
D	1285	CGTCCCAATTCCCGGGTCCCGGCGACCGGTGTGTATGAGTGTCCAAACCGGGTCTGT	1226
Q	1624	gaacgcctctcaagtcgcgtgtcgcgcggaacgaagcgcaaggtctgcgcgggtlaaagaatcccg	1683
D	1225	GACGCCCTTGGGTCCGGTGTGGCGCGGAGAACGAGCGAGAGGTGCGACGTTAAACACACAG	1166
Q	1684	ggcggcccgatgtcgtctcgtgtcaacagctcgaaacccgacggacggacggcgacgcgcgt	1743
D	1165	GGCGGTCCGAGTGTGTCGTGATCACCAGTTCACCCAGCGACCGAGCGGGCACGCCGT	1106
Q	1744	caaccttcggacctcatagacctccctctgcgcgaagccgcgaaacggcgcccgcggtgtc	1803
D	1105	CACCTTCGGGCTCTCATATGACCCCTTGTGGCCACACCCCGGAACGGGGCCCCCGATGGGTTC	1046
Q	1804	ggcgaagaacgcacatccgcgcggaagcggaagcgtlcaatcggatgcagatgtaactctgc	1863
D	1045	GGCGAAGACCGGCATTCGGGCCAGACCGAGACGTTCCATCGGGGTCTATGATATACGTCCGC	986
Q	1864	ggccgcgcgtctcgaaacgcgtctgttcaagtggtctgcgcgtctcgtgtgcgtgcaagttaacggt	1923
D	985	ACCGCGCGTCTGACACCCGTTTGTGTAGTGCATGTCGCGCGTGTGTTGGTGGGAATACAGGT	926
Q	1924	ccaagatgacggcgaatgcgtgtctgttcgtctccgcggccggtacggcggtgtgtaggggt	1983
D	925	CCAGATGCGCGGCAATGCCCTGTGTGTGTTGTCGCGGGCCCTGTACGGAAGGTGTGTAAGCGGT	866
Q	1984	gtcgaatctgtgtggcggggaacggcggaacagctctccgttcgagagctggaagctgtgta	2043
D	865	GTTCGATCTATGCGGGGCAACCGCGCGACGCGACACAGCTTCCGTTCCGTTGGAGCTGAAGTGTGTA	806
Q	2044	tcaccgcggcccggtctcgtctgaacacaggttgacggttccagccgaacagagcggtgtlaaga	2103
D	805	TCGCCCGGACCCGGGTGCTGTACCAACAGGTGCGGTTCAGCCGAAGAGGCGCGTGTAGAA	746
Q	2104	gtcgcgcgaagcgcgcaacatcgggtcgaaacgaagtctgaacaatgcgggggtgcgcgcgc	2163
D	745	GTTCGCGCGAGCGCGGACACTTCGGGCGACCCGAGGTGTGAACACTTCGCGGGGCGCCCGCGCGC	686

[illegible][illegible]

QY	1084	gctccgaacgaggagatctcgaacccgtctcaacccccccgaagcggcgggagagat	1143
Db	1765	gctccggaaaggagatgtcgaagacctcgtccaccccccccgaagcggtagat	1706
QY	1144	cggggtccctcaacgaacccgggagatcggcagcaacctcggcggaaggctcaatcgttga	1203
Db	1705	cggggtcccttcaagaaacggggagatcgagccagcaacctcgcgggcgagggcgatcatggtga	1646
QY	1204	tgaaaccaaggagggttcggccggcgggggggttcaactcgaacgaacgggcgcgtccgga	1253
Db	1645	tgaccccaagggggtccggccggcgggggcggttcaactcgaacgaacggggcgccgctctcga	1586
QY	1264	aggcgtctgacactcggggtcttcgtctgatgtcaatacctggaacgcgaacggtgtga	1323
Db	1585	atcgctgtgtcactcactcggggttccgctgtcatctatctatggagaaacgaagaggggtga	1526
QY	1324	cggcaacccggggcgccctcgccgtcaagaaacacccgcgcgcgcgggttaacctc	1383
Db	1525	cggcaacccccggcgccctccctctatgaagcaccgtccgcgcgggggttagccttc	1466
QY	1384	ccggaagtctgaacacacgaagaagacggcccgctggcggtctgagaacgaatctcgcccgagcc	1443
Db	1465	ccggaagtccaacacacgaagaccggcccgcccgctggcggtctgagaacgaatctcgcccgagcc	1406
QY	1444	ccctggcggtctcgcgggggtcgtctctcgaacgaagccggaattcaacggcgcgccgagac	1503
Db	1405	ccctggggggtgtcacagggggctcgtttcgaacgaagccggcgaattgacagggcccgccgagac	1346
QY	1504	cgctcggtcgcaatctctgaacggcggaagacacggcggaagtcggcggtacacggcggt	1553
Db	1345	cgctcggtcccaatctctgaacggcggaagacacggcggaagtcggcggtacacggcggt	1286
QY	1564	gaaccccaagctcctctgggttgccggcggaacgtgtgtcaacgaatgtccagacccgggtcgt	1623
Db	1285	cgctcccaagtcccggggtgcccggcggaacccgctggtgcatgattgattgccaaacggggtcgt	1226
QY	1624	gaaccccttaagtctcgggtgtctcggcggaacgaagccggaaggtctcgccgggtagaagtcccg	1663
Db	1225	gaaccccttccgtccggtgtcggcggaacccggaacccggaaggtctcgccgggtagaagtcccg	1166
QY	1684	ggcgcccgcatccggttcggttcaacgaagctccgaacccgaacccggcggaacccggt	1743
Db	1165	ggcgcggtcccatgtcgtcgtgtcacgaattccacaccagccagacacggcgagccggt	1106
QY	1744	caectccgagccctcaatgaactcctcttgcgcgaagacggcgaaacggcccgcggtctc	1803
Db	1105	caactccgagccctccatgaacgaccttggcgcaacaccggaacggcgagcccgagtggtc	1046
QY	1804	ggcggaagaccgcatcccgccggcggaaggaacgttcaatcgggaatcatatgaatgaactgc	1863
Db	1045	ggcggaagaccgcatcccgccggcggaaggaacgttcaatcgggaatcatatgaatgaactgc	986
QY	1864	ggccgcgcgtctgaacccggttgaatgaatgcgttcggcggtctggttggcggaagtaacga	1923
Db	985	accgcgcggttcgacccggttgggtcagtgccgtgcggtgcgttggcggaagtaacga	926
QY	1924	ccgaatggccggcatgcgtcgtcgtcgttctccggggccgtgaacggcggttgtaggggt	1983
Db	925	ccgaatggccggcatgcgtcgtcgtcgttctccggggccgtgaacggcggttgtaggggt	866
QY	1984	gtcgaatctgttggcggaacccggcggaacgaactcccgctcgagagctgaacgtctgtga	2043
Db	865	gtcgaatctgttggcggaacccggcggaacgaactcccgctcgagagctgaacgtctgtga	806
QY	2044	tcccccggcgccgggtctcgttgaacgaagttggtcgttcaacggccaagacgaagcgggtgtagaa	2103
Db	805	tccccccggcacccggggtcgttgaacgaacgggtcgggtccacgggaagagccggttagaa	746
QY	2104	gtcgccgaaggcggaacatctcggcggaacgaaggttgaaacaaatgctggggggcgccggcggc	2166
Db	745	gtcgccgaaggcggaacatctcggcggaacccgaagtggaacaaatgctggggggcgccggcggc	686

[illegible]

Query Match	88.6%;	Score 2542.2;	DB 6;	Length 3196;
Best Local Similarity	93.6%;	Pred. No. 1.3e-202;		
Matches 2681; Conservative	0;	Mismatches 163;	Indels 21;	Gaps 2;

OY	5	cggaaeccggtgatacaacgycgaacgycgcgccacaaccgcyggttgcgcyggttggtccggt	64
Dd	2845	CCGCAACCGGGGAGACC GGCCAGCGGGACGCC CACC GCCCGATGGGGCGGTGCCGTCCCGTG	2786
OY	65	cc-gatcgcgcgcgcgcgcgatcaagcgcaacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	123
Dd	2785	CCGGGTCCGGGTCCGGCCCATTCACGCACGCCAGACGGGGCAGTCACTGAACCCGCGCTTGCT	2728
OY	124	ggcccccttcgcgcaccaccgcgaactcgttcgtacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	183
Dd	2725	GGGGCCCCCTTCCTCGCACACCGCAACTCCTCTGACGGCACGGCGCATCGGCTTGGGGGA	2666
OY	184	aactgctgcygafragaacgcgatatacgttgcgcgcaccccgcgcgcgcgcgcgcgcgcgcgcgcgc	243
Dd	2665	AACTGCTCGCAATGACCCGATATCTGTGGCATTTCCAGCTGGGGAACTTCCTCCCGA	2606
OY	244	tgcagtaagtgc	303
Dd	2605	TGCAGTAATGCGGGCCGCTGCCCGAAGGTAGCGCCGCCCTCGAAGGCGCGTCCGGGTGGA	2546
OY	304	aggcgtcgc	363
Dd	2545	AAGCGTGGCGGGCTCGCTGATGGCGGGCGCTCGGTGTTGGCCCTCGATGCACACAGTA	2488
OY	364	ccgc	423
Dd	2485	CCGGCGCTCCGGGGGACCGCGGACACCGCGGATGACCACTCCGCGAGCGAACTTC	2428
OY	424	acaaagctgtagaacacgc	483
Dd	2425	ACAAGAATAGCGCACCGGGGGGTGGCGCAGCGCTCTCTCAAGAACCGGAGATGG	2366
OY	444	cgtctcgttcgc	543
Dd	2365	CGTCCCGCTCGGCTTCGCCCGCGCAGACGGGCCCCACGAGCACTCCGCAAGCAAGGCCCA	2306
OY	544	ggaaggaacgc	603
Dd	2305	GGAAGAACCGGTGTCTCGTGGCCGGCGGAAATGAGTCCGTGATCATGTAGACGACT	2246
OY	604	ggtcgttcgc	663
Dd	2245	GGTGTGCGAACCAGAACCCCAACTCTGCTGTCGCTTCAATAGACACGGCGGTCAATGG	2186
OY	664	tcggggtggttcgc	723
Dd	2185	TCCGGGTCTCTTCCGCGCGGAGCGCGTGACGGCTTCGAGGAAAGGCTTCCACCGCGCG	2126
OY	724	aggtgtccgc	783
Dd	2125	ACGTGTCCGGACGCTCCCGGACGGGTCTCGTGGCTCACCGCGCGGCTCTCGGGGCGCG	2066
OY	784	cgaagccgaagtgccttgagaaacgcctgaacgc	843
Dd	2065	CGAAGCCGAGAGTCTTGAAGCACCCGACGGCGCTCGGGGCGCATTTGCCGGAATGATGACCG	2006
OY	844	gcacacccgc	903
Dd	2005	GCACGCCGAGATTGCCAATATACCAACACGCGGAAAGTGATACCGAAGCGCGCGATCA	1946
OY	904	gctcgc	963
Dd	1945	GCTCGCGCGGTTCCCGGACCGGTCGAGAGTAGTGGCGAGTTCGCTGAGACACCGCTGGG	1886
OY	964	cgatcgc	1023
Dd	1885	CGATGCGCGGATCCGATCCGTCCTCGCTCGGCGCACGCGCGCGGTGAACCGCGGTGCGT	1826
OY	1024	ggatgc	1083
Dd	1825	GGATGCGCGCACGCGCGGTTGCTCTCAACCGTTCACGCGCGATGAGCGTGAACGGGACGCCA	1766

[illegible]

JOURNAL MICROBIOL BIOTECHNOL (2001) In press
REFERENCE Z (bases 1 to 1932)
AUTHORS Hong,Y.-S. and Lee,J.J.
TITLE Direct Submission
JOURNAL Submitted (26-JUL-2001) Anticancer Research Laboratory, Korea
Research Institute of Bioscience and Biotechnology, Yulsong, P. O.
Box 115, Taejeon 305-600, Korea
location/Qualifiers

[illegible]

```

||||| 1276
Db 1335 GCTGTTGTGAGAGACCGACCGAATCGGCTGCGTGTCTGTACAGCACCGGGTCA 1276
Oy 661 tggatcgaggtggttccggcgaggttgatgcaaggttcagagagcagctctccagg 720
Db 1275 TGGTGGGGGTGTCTTCGCGGGGCTGAGTGCAGGGCTTGAGAGAGCAGGCTCTCCAGG 1216
Oy 721 ccgaggtgtctcgcagacgcccccgaggtctgtccgtcaaccccccgagctctggggc 780
Db 1215 CCGAGGTGTCCGGACCGCCCCGGCAGGGTCCGTGCGGTACCCCCCGCTCT--GGC 1158
Oy 781 ccgcagagcgagatgctcttgagaaagctgcaaggttcggcgagcatcgcgagatcggtga 840
Db 1157 CGCCGAGG-CGAGTGTCTTGAGAAAGCTGACGGCTCGCGGGCCAT---CGGAACGGTGA 1102
Oy 841 ccgagacacgagacgagctcgagatgatacaacagcggaagtggtacggaagcgcgcga 900
Db 1101 CCGGCACACCGAGCAGCTCGAGATGACCAACAGCGGAGTGTACGCAAGCCGCCGA 1042
Oy 901 tcaagtcgaggttctgcccagacgagcgagagcggtcgaggttcgttgagcagcgagc 960
Db 1041 TCAGCTCGGCGCGTTCGCCCGCACCGCGGAGGCTCGCGCAGTTGCGTGAAGCACCGGCG 982
Oy 961 ccgagatcgagcgagatgagatcgctcgcgtcgagcgagcgaggttgagacgagtg 1020
Db 981 CGCGCATCCGCGGATGCGATCCGTCCGCTCGGCCAGCGCGGGGCTTGAAACGAGGTG 922
Oy 1021 cgtgatgagcgagcgagcgaggtgagctcgctccagcgagatgagcgttgagcagac 1080
Db 921 CCGTGTGATGGGATGAGAGCGCGCGGTGGGCTCGCCGTCACGCGGATGAGACGAGGAC 862
Oy 1081 gcaagctcgagaaagggatgtcgagacgctgtccaccccccgagcgagcgagggcgga 1140
Db 861 GCAGCTCCGGGACGGGAGTGTGAGAGCCCTGCAACACCCCGCCAGCGGGGGGGCGGA 802
Oy 1141 ggtcgaggtctctcagcaacgagggatcgcaagcagctcgagcgagggcggtacatcg 1200
Db 801 GGTGGGGTCTTTCACCAACCGGGGATCGGCCAGACATAGCGGAGAGGCGTCAATCGG 742
Oy 1201 tgaatgacccagcgaggtccgcccggagggcggtcactcagacagcagcgagcggtccc 1260
Db 741 TGATGAACCTAGGCGGGTCCCGCAGAGGGGCGTTCACCTCAACGAGCGGGCGCTCCG 682
Oy 1261 ggaaggtgtgtgactcagcgaggttcgagtgatcatcagggagacagcgagcggt 1320
Db 681 GGAAGGCGTGTGACTCGGCTTGCGCTGCAATGCTCATGAGGACAGCGAAGCGGT 622
Oy 1321 cgaagcgacacgagggcgagctcgagctcagaggaac--gcccgcgcgcgcgcgcgcgcgc 1379
Db 621 CGACGGGACACCGGGGGCGCTCGCGCTCAGAGGCACTGCGCGCGCGCGGGGTACC 562
Oy 1380 cctcccgaggttcgacacagagagagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1438
Db 561 CCTCCCGCAGTTCCAGCACCGAGAGAGCGGCCCGCTGACGGGTGAGACAGTCCGCCAC 502
Oy 1439 cgcgcgccttgagcgt--gtcgagcgagcgtctctcagcagagcgagcgaggttcagagcgcg 1497
Db 501 CGCCCGCTGGGCGTATGCGGGGCTGTTCTCAGCGAGGCGCGGATTCAGCGGCGG 442
Oy 1498 ccgagacgctcgagctcgagctgtgcaagcgagaaagcagcgagcgaggtcgagcgagcg 1557
Db 441 CCGGACCTCGCGTGCAGTGTGTGACGCGGACAGCAGCGCCCAAGTCCGCGCTTACGCG 382
Oy 1558 gccggtgagcagcagcagctcctgtgtgtcgagcgagcgaggtgtgtcacagatgttcagagcgcg 1617
Db 381 GCGGCTGACGCCAGCTCTGGGTGCCCGCAGCGGTGTGTGACCGGATGTCCAGACCGG 322
Oy 1618 gtcaggtgagcagcagctcagctcagctgtcgagcgagcgagcgaggtcgagcgagcg 1677
Db 321 GTCAATGAGCGCTTCACTGCGGTGTCCGCGGAGCAGGCGCGAGGTCGCGGGGTAGAA 262
Oy 1678 gtcccgagcgagcagcgagctcgagtcagcagcagcagcgagcgagcgagcgagcgagc 1737
||||| 1737

```

```

Db 261 GTCCCGGGCGGCCCGCATGCCGTGCTACCAAGTTCGACCCAGCCGAGCCGAGCGGGCAC 202
Oy 1738 gccggtcagctcgcgcgcctccatgactcctctgtgcccagacgagcgagcgagcgagc 1797
Db 201 GCCCGTACCTCCGCCCTCCATGACTCCCTTGCGCCAGAGACCGAAGCGGGCG---GG 145
Oy 1798 ggggtcgagaaagccgcacatccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1857
Db 144 CGGGTGGCGGAAGCCGATCCGCGCGAGGCGCGAGGACGTCATCGAGTATGATGAC 85
Oy 1858 ctgcgcgc--ccgcgcgcctcagacccgcttgatgagtgagtgagtgagtgagtgagtgag 1916
Db 84 CTCGCGCGCCCGCGCTCTGATCCAGCCGCTTGCTGATGTCGTGCGGTGCGGAGT 25
Oy 1917 acaggtcagatgagcgagatgc 1940
Db 24 ACAAGTCTAGATGAGCGCGCATGC 1

RESULT 6
184584/c 1269 bp DNA linear PAT 04-APR-1998
LOCUS 184584
DEFINITION Sequence 1 from patent US 5695966.
ACCESSION 184584
VERSION 184584.1 GI:3022104
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1269)
AUTHORS Inventi,A.Solari., Breme,U., Colombo,A.Luisa.,
Hutchinson,C.Richard., Otten,S. and Scotti,C.
TITLE DNA encoding daunorubicin 14'-hydroxylase and method for preparing
doxorubicin
PATENT Patent: US 5695966-A 1 09-DEC-1997;
JOURNAL Location/Qualifiers
FEATURES
1..1269
source /organism="unknown"
BASE COUNT 164 a 484 c 442 g 179 t
ORIGIN

Query Match 44.2%; Score 1269; DB 6; Length 1269;
Best Local Similarity 100.0%; Pred. No. 4,5e-97;
Matches 1269; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 83 tgaagcgagcagagcgaggttcggtgagcgagcgagcttcggccccctccgcagca 142
Db 1269 TCAGCGCAGCCAGACGGGCGATTCGCTGAGCGCGCGCTGTGGGCCCTTCGCGCACCA 1210
Oy 143 ccgcaactgtgtgagagagcgagcgagctcgagcgagcgagcgagcgagcgagcgagcgagcg 202
Db 1209 CCCCAACTGTGTAGGAGAGCGGCGCAGTGGGCGCTCGGGGAACCTGCTCGCATACGCGC 1150
Oy 203 gatcatcgtgcgagctcagctgagcgagctgtcctccagatgagtagtcgagcgagcg 262
Db 1149 GATCATCTGTCGGGACTCCAGCTGGGCGAGTCTCCCGATGCAATGAGGCGCGCTG 1090
Oy 263 gccgaagtgagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 322
Db 1089 GCCGAAGTGAAGCGCGCGCGCACGAGGAGCGTCCGGGTGGAAGCGCTGGGGCGCTG 1030
Oy 323 atgagcgagcgagtggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 382
Db 1029 ATGGCGGCGCTGCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 970
Oy 383 ccgagacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 442
Db 969 CCGAGCGCGCGCGATGTACCTCCGTGGCAGCGAACCCTCAACAGCTGTAGGCAACCGG 910
Oy 443 cgggtgtgagcgagcgagctcctccagaaacgagcgagcgagcgagcgagcgagcgagcgagcg 502
Db 909 CGGCTGTAGCGCAGCGCTCTCTCCACAGACCGGGAAGCGGCGTCTCTGTCGCGCATCCG 850

```

LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	ORGANISM	REFERENCE
AR078148/c	Sequence 7 from patent US 5962293.	AR078148	AR078148.1	GI:10004894	Unknown.	1 (bases 1 to 2081)
LOCUS	Sequence 7 from patent US 5962293.	AR078148	AR078148.1	GI:10004894	Unknown.	1 (bases 1 to 2081)
DEFINITION	Sequence 7 from patent US 5962293.	AR078148	AR078148.1	GI:10004894	Unknown.	1 (bases 1 to 2081)
ACCESSION	Sequence 7 from patent US 5962293.	AR078148	AR078148.1	GI:10004894	Unknown.	1 (bases 1 to 2081)
VERSION	Sequence 7 from patent US 5962293.	AR078148	AR078148.1	GI:10004894	Unknown.	1 (bases 1 to 2081)
KEYWORDS	Sequence 7 from patent US 5962293.	AR078148	AR078148.1	GI:10004894	Unknown.	1 (bases 1 to 2081)
ORGANISM	Sequence 7 from patent US 5962293.	AR078148	AR078148.1	GI:10004894	Unknown.	1 (bases 1 to 2081)
REFERENCE	Sequence 7 from patent US 5962293.	AR078148	AR078148.1	GI:10004894	Unknown.	1 (bases 1 to 2081)

[illegible]

[illegible]

Search completed: June 11, 2002, 20:01:13
Job time: 7335 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 11, 2002, 17:59:13 ; Search time 308.46 Seconds

(Without alignments)
15974.663 Million cell updates/sec

Title: US-09-673-254-1

Perfect score: 2870

Sequence: 1 ggcacgcacgcgcgtacacg.....catcgatgcgcgcgtacacg 2870

g table: IDENTITY-NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 segs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2870	100.0	2870	AA231413	Complementary DNA
2	2558.2	89.1	3196	AAV01453	Streptomycetes dauno
3	2342.2	88.6	3196	AAV01447	Daunomycin C-14 hy
4	1269	44.2	1269	AAV013617	DNA gene encoding
5	1224.8	42.7	3013	AAV01451	Plasmid pANT195 in
6	1221.6	42.6	2081	AAV01452	Modified doxa gene
7	1044.4	36.4	1569	AAV34001	DNA fragment conta
8	182	6.3	14955	AAV33491	Human adenosine Al
9	181.2	6.3	125401	AAV17186	Streptomycetes nous

C 10	164.4	5.7	114955	20	AAV33491	Human adenosine Al
C 11	154.8	5.4	2919	21	AAZ24486	Streptomycetes tende
C 12	152.6	5.3	12249	21	AAV55840	Complete Mycomycin
C 13	152.6	5.3	18331	21	AAV55857	Complete nucleotid
C 14	152.2	5.3	2918	21	AAZ43922	S. tendae nikomyc
C 15	142	4.9	65140	22	AAV17184	Streptomycetes nous
C 16	139.8	4.9	1257	22	AAV81357	Quorum sensing con
C 17	138.6	4.8	1224	21	AAV5788	Hydroxylase orf4.1
C 18	137.4	4.8	4257	19	AAV68520	The nucleotide seq
C 19	137.4	4.8	4257	19	AAV68520	Infected cell prot
C 20	137.4	4.8	5970	21	AAV15635	Nucleotide sequenc
C 21	137.4	4.8	5970	21	AAV56003	Contig 002 from co
C 22	133.8	4.7	1251	21	AAZ87301	S. venezuelae macr
C 23	133.2	4.6	4496	17	AAV58553	Streptomycetes prist
C 24	132.6	4.6	12001	16	AAV6213	HSV L/SF region.
C 25	130.6	4.6	44377	18	AAV8508	Platenolide syntha
C 26	130.6	4.6	44377	18	AAV80414	Sequence soyC and
C 27	128.6	4.5	1700	14	AAQ45569	Streptomycetes grise
C 28	127.4	4.4	2795	22	AAV11296	Streptomycetes grise
C 29	127.4	4.4	2795	22	AAV81954	Streptomycetes grise
C 30	127.2	4.4	27541	22	AAV17185	Streptomycetes nous
C 31	123.8	4.3	3957	22	AAV09686	HSV-2 immediate ea
C 32	123.8	4.3	154746	24	AAV25519	Human herpesvirus
C 33	123.8	4.3	154746	24	AAV25519	Human herpesvirus
C 34	122.2	4.3	109519	22	AAV08653	Micromonospora DNA
C 35	122.2	4.3	58857	21	AAV58471	Nucleotide sequenc
C 36	121.6	4.2	1194	17	AAV58555	Streptomycetes prist
C 37	120.8	4.2	14806	22	AAV03809	Streptomycetes galli
C 38	119	4.1	940	19	AAV63943	Mycoacterium tube
C 39	119	4.1	940	20	AAV81092	Nucleotide sequenc
C 40	117.6	4.1	24379	18	AAV3095	Streptomycetes fireo
C 41	117.6	4.1	24379	19	AAV25925	Streptomycetes roseo
C 42	117	4.1	50937	21	AAV0469	Streptococcus olea
C 43	115.6	4.0	8438	15	AAV73500	DNA encoding pseud
C 44	114.6	4.0	1998	12	AAQ11127	Sequence encoding
C 45	113.8	4.0	1879	12	AAQ11126	Sequence encoding

ALIGNMENTS

RESULT 1	
AA231413	
ID AA231413 standard; DNA; 2870 BP.	
XX	
AC AA231413:	
XX	
DT 07-FEB-2000 (first entry)	
XX	
DE Complementary DNA consisting of doxa, dnrv and C-terminal dnrv genes.	
XX	
KW doxa gene; daunorubicin 14-hydroxylase; daunorubicin; doxorubicin;	
KW anticancer agent; dnrv gene; dnrv gene; complementary; ss.	
XX	
OS Streptomycetes peucetius.	
XX	
PN NO9955829-AA2	
XX	
PD 04-NOV-1999.	
XX	
PF 22-APR-1999; 99WO-US07016.	
XX	
PR 24-APR-1998; 98US-0065606.	
XX	
PA (PHAA) PHARMACIA & UPJOHN SPA.	
XX	
PI Solari A, Zanuso G, Filippini S, Torti F, Otten S, Colombo AL;	
PI mutacin CR;	
XX	
DR WPI: 2000-023353/02.	
XX	
PT New DNA containing the gene for daunorubicin-14-hydroxylase and genes	
PT for resistance to anthracyclines, used to prepare the anticancer agent	

[illegible]

RESULT	2
AAV01453/c	
ID	AAV01453 standard; DNA; 3196 BP
xx	
AC	AAV01453;

21-MAY-1998 (first entry)

Streptomycetes daunomycin biosynthesis gene cluster.

Daunomycin C-14 hydroxylase: doxa gene; doxorubicin; daunomycin; 13-dihydrodaunomycin; carminomycin; anthracycline; anticancer; cytostatic; cancer; therapy; ds.

Streptomycetes sp. strain C5.

Key Location/Qualifiers

CDS 9..590

FT /tag= a

FT /note= "3' end of orf1 coding region"

FT 663..666

FT /tag= b

FT 674..1501

FT /tag= c

FT /product= daunomycin C-14 hydroxylase

FT /note= "orfA"

FT 1498..2766

FT /tag= d

FT /note= "doxa"

FT 3038..3041

FT /tag= e

FT 3050..3196

FT /tag= f

FT /note= "5' end of dau1 coding region"

CDS

RBS 3038..3041

RBS 3050..3196

22-MAY-1997; 97WO-US08690.

24-MAY-1996; ABUS-0653650.

(OHS) UNIV OHIO STATE RES FOUND.

Desanti C, Dickens ML, Strohl WA.

WPI: 1998-018495/02.

P-PSDB; AAW36128, AAW36130 AND AAW36133-34.

Producing doxorubicin from daunomycin by daunomycin C14 hydroxylase - also hydroxylation and oxidation of other anthracycline(s) with the same enzyme

Disclosure: Fig 3; 59pp; English.

This DNA sequence comprises a 3196 bp KpnI-SstI DNA fragment of Streptomycetes sp. strain C5 containing the doxa gene (see also AAW01447) that codes for daunomycin C-14 hydroxylase (see AAW36128). The DNA fragment also includes the 3' end of orf1 (see also AAW36133), the complete orfA (see AAW36134) and the 5' end of dau1 (see 36130), a putative transcriptional activator. Daunomycin C-14 hydroxylase is an enzyme capable of converting daunomycin to the anticancer agent doxorubicin. Host cells, especially Streptomycetes host cells, transformed with plasmids that include the doxa gene can be used in methods for the production of doxorubicin from daunomycin or for the hydroxylation and oxidation of other anthracyclines.

Sequence 3196 BP: 475 A: 1182 C: 1078 G: 461 T: 0 other;

	Query Match	89.1%	Score 2558.2;	DB 19;	Length 3196;
	Best Local Similarity	93.98%;	Pred. NO. 0;		
	Matches 2691; Conservative	0;	Mismatches 153;	Indels 21;	Gaps 2;
QY	5 ccgcaccggtctacacagcagcgagcacccgccgcgcgctgcgggttgaggcgtcccgctg	64			

[illegible][illegible]

[illegible]

24-MAY-1996: 96US-065365D
(OHIS) UNIV OHIO STATE RES FOUND.
Desanti CL, Dickens ML, Strohl WA
MPI: 1998-018495/02.
P-PSDB; AAW36128.
Producing doxorubicin from daunomycin by daunomycin C14 hydroxylase
PT - also hydroxylation and oxidation of other anthracycline(s) with
the same enzyme
Claim 11, Page 30-33; 59pp: English.
This DNA sequence comprises the doxa gene of Streptomyces sp.
strain C5 that codes for daunomycin C-14 hydroxylase (DC14H)
(see AAW36128), a P450-like enzyme useful for the production of
doxorubicin from daunomycin, and for the hydroxylation and
oxidation of other anthracyclines. The gene was identified by
sequencing an approximately 8 kb fragment of the daunomycin
biosynthesis gene cluster from C5 (see AAV01453). The gene has been
incorporated into various vectors, including pANT195 (see AAV01451)
and pANT199 (see AAV01452). Claimed methods for production of
doxorubicin include: the use of a host microorganism transformed
with a plasmid which contains the doxa gene; incubating a culture
of Streptomyces sp. strain C5 with daunomycin; and incubating
DC14H with daunomycin. The enzyme can also be used in claimed
methods of producing 13-dihydrodoxorubicin and carminomycin
from 13-deoxydoxorubicin, and of producing 13-dihydrodaunomycin
and daunomycin from 13-deoxydaunomycin. The methods preclude the
use of halogens as required in the current chemical methods.

Query Match	Similarity	88.6%	Score 2542.2	DB 19	Length 3196
Best Local	Similarity	93.6%	Pred. No. 0		
Matches 2681	Conservative	0	Mismatches 163	Indels 21	Gaps 2
OY	5	ccgcacccgggtacacagcagcagcagccacccacccgcgcgtgtgcgcgttgcgcgcgtg	64		
Db	2845	CCGCACCGGGGGGACCGGGACCGGGACCGGCCCAACCGGGCATGGCGCGTGGCGCGTCCG	2786		
OY	65	cc-ggtcgc	123		
Db	2785	CCGGGTCCCGGTGGCGCCCATCAACGACGACGACGAGCGAGGTCACTGAGCCCGCTGTCT	2726		
OY	124	gggcccccttcgc	183		
Db	2725	GGGCCCCCTTCCTCGACACCGCAACTCTCTGTAGGACGGCCAGTCGGGCTTGGGGGA	2666		
OY	184	acctctcgc	243		
Db	2665	ACCTCTCTCGCAGTACGCCGATCATCTGCGGATTCCAGCTGGGCGAGCTCTCCCGA	2606		
OY	244	tgacgtatgc	303		
Db	2605	TGCATTAATGGGCGCCCGTGGCGAAGTGAGCGCCGCGCTGSAAGGGCGTCCGGGTGA	2546		
OY	304	aggcgtgc	363		
Db	2545	AAGCCTTGGGGGCGCTGTGATGCGCGGCGGTGGTGTGTGATGCCCTTCGATGTCACACAGA	2486		
OY	364	ccgc	423		
Db	2485	CCGGGCGTCCCGGGGGGACGGGAGACCGCGGATACACCTTCCTGTGACAGGAACCTCC	2426		
OY	424	acaaactgttaagc	483		
Db	2425	ACAACGATTAAGGCGACCGGGCGGTGTGTGGCGAGCGCTCTCTCCACGAACGGGAGATGG	2366		
OY	484	gctcctgc	543		

[illegible][illegible]

Db	129	CGCGGGGCGCTTACACTGCAGACCGGCGCCGCTCCCGAAGGCGCTGTCGACCTCG66	70
Oy	1283	cttgcgtcgaatgctcatcatcggaacgcggaaacgggttcgaagccaccgggcgcctc	1342
Dd	69	CTTGCGCTGCATGTCAATCANGGACACGCCGAACGGGATGACGCGCACCCGGGCGCTC	10
Oy	1343	gccgcctcac	1351
Db	9	GCGCGCTCAC	1
 RESULT 5			
ID	AAV01451/C		
XX	AAV01451 standard; DNA; 3013 BP.		
XX	AAV01451;		
XX	21-MAY-1998 (first entry)		
XX	Plasmid pANT195 including the doxa gene.		
KM	Daunomycin C-14 hydroxylase: doxa gene: doxorubicin: daunomycin:		
OS	13-dihydrocarbamoylcinnamic acid: carbinomycin; anthracene;		
XX	anticancer; cytotoxic; cancer; therapy; Plasmid pANT195; ds.		
XX	Streptomyces sp. strain C5.		
FH	Key	Location/Qualifiers	
FT	CDS	complement (99..1034)	
FT		/*tag= a	
FT	-10_signal	/note= "snpr activator gene"	
FT		complement (1043..1047)	
FT	-35_signal	/*tag= b	
FT		complement (1067..1071)	
FT	protein_bind	/*tag= c	
FT		1193..1209	
FT		/*tag= d	
FT	-35_signal	/note= "proposed snpr binding site"	
FT		1230..1235	
FT	-10_signal	/*tag= e	
FT		1252..1257	
FT	RBS	/*tag= f	
FT		1262..1265	
FT		/*tag= g	
FT	RBS	1302..1305	
FT		/*tag= h	
FT	CDS	/note= "engineered RBS site"	
FT		1315..2583	
FT		/*tag= i	
FT	RBS	/note= "doxa gene"	
FT		2855..2858	
FT		/*tag= j	
FT		2867..3013	
FT		/*tag= k	
FT		/note= "5' end of daul gene"	
PR	M09744439-A2.		
PD	27-NOV-1997.		
PE	22-MAY-1997; 97MO-US08690.		
PR	24-MAY-1996; 96US-0653650.		
PA	(OHIS) UNIV OHIO STATE RES FOUND.		
PI	Desanti CL, Dickens ML, Strohl WA;		
DR	WPL: 1998-018495/02.		
DR	P-PDB: AAM36128-30.		
PT	Producing doxorubicin from daunomycin by daunomycin C14 hydroxylase		

PT	- also hydroxylation and oxidation of other anthraacyclines) with
PR	the same enzyme
XX	
PS	Disclosure: Fig 6; 59pp; English.
XX	
CC	This DNA sequence comprises a fragment of plasmid pANT195
CC	containing the Streptomyces snpr activator gene, the snpr-
CC	activated promoter and a 5' end-modified (added ribosome binding
CC	site) doxa gene (see also AY01447) derived from Streptomyces sp.
CC	strain C5. The doxa gene codes for daunomycin C-14 hydroxylase
CC	(see AA036128), an enzyme capable of converting daunomycin to
CC	the anticancer agent doxorubicin. Host cells, especially
CC	Streptomyces host cells, transformed with pANT195 can be used
CC	in methods for the production of doxorubicin from daunomycin or for
CC	the hydroxylation and oxidation of other anthraacyclines.
XX	
SQ	Sequence 3013 BP; 448 A; 1074 C; 1066 G; 425 T; 0 other;

[illegible]

Db 1942 ACGTGTCCGGCAGTCCCGGCGAGGCTCGTGCCTACACCGCCGCGCTCTCCGGGCGCC 1883
 Oy 784 cgaagccagatgctctgagaacagctgagcgctcgaggccatcgccgagtcgtagccg 843
 Db 1882 CGAGGCCAGTCTCTTGAACACGCGGAGCGCTCGCGGGCCATTGCGGATGCTGACCG 1823
 Oy 844 gcaacacagcagcagcagatgacacacagcagcagcagcagcagcagcagcagcagcagc 903
 Db 1822 GCACCCCGAGGAGTTCGAGATGACCAACAGGGGAAAGTGTACCGCAAGCCCGCATTA 1763
 Oy 904 gctcgccgctgttcgcaacagcagcagcagcagcagcagcagcagcagcagcagcagc 963
 Db 1762 GCTCGCGCGCTTCCGCGACCGCTGCGAGAGTTCGATGAGACACCGGCTCG 1703
 964 cgaatcgaggagatgcatccgctcgctcgccgagcagcagcagcagcagcagcagcagcagc 1023
 1702 CGATGCGCGCGAGTGCATGCTCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 1643
 Oy 1024 ggaatcgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1083
 Db 1642 GGATCGCGCGAGAGCGGGGTGCTCTCACCGCTCACGCGCGCATGAGCGTGAACGAGCGA 1583
 Oy 1084 gctcgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1143
 Db 1582 GCTCGGGAACGGGGATGTCGAGACCGCTGCTCACCGCGCGCGCGCGCGCGCGCGAGAT 1523
 Oy 1144 cggagctctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1203
 Db 1522 CGGGTCTCTTCAGAACCGGGATGCGCGAGCATCTCGGGCGAGGGGTATGAGTGA 1463
 Oy 1204 tgaacacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1263
 Db 1462 TGACCCAGCGGGGTGCTCGCGCGCGGGGGGCTTACCTCGACGACGAGGGCGCCGCTCTCGGA 1403
 Oy 1264 aggcgtctgacccctcgagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1323
 Db 1402 ATGCTCTGTCGACCTCGGGGTGCTCGCTCATGCTCATGAGGACAGAGAGGCTCGA 1343
 Oy 1342 cggcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1363
 Db 1342 CGGCCACCGCGCGCGCTCGCGCGCTCATGAGGACACCCCTC 1303

RESULT 6
 AAV01452/C
 AAV01452 standard; DNA; 2081 BP.

AAV01452:
 21-MAY-1998 (first entry)
 Modified doxa gene encoding daunomycin C-14 hydroxylase.

XX DE 21-MAY-1998 (first entry)
 XX XX Modified doxa gene encoding daunomycin C-14 hydroxylase.
 KW Daunomycin C-14 hydroxylase; doxa gene; doxorubicin; daunomycin;
 KW 13- α -hydroxycarmine; carminomycin; anthracycline;
 KW anticancer; cytotoxic; cancer; therapy; plasmid PANT199; ds.
 XX OS Chimeric - Streptomyces sp. strain C5.
 OS OS Chimeric - synthetic.
 XX XX

Key Location/Qualifiers
 FH -35_signal 3..8
 FT /*tag= a 26..31
 FT -10_signal 38..58
 FT /*tag= b 38..58
 FT /*tag= c 74..142
 FT /*note= "lac operator"
 FT misc_signal 74..142
 FT /*tag= d
 FT /*note= "rrnb antiterminalator"
 FT enhancer 159..168
 FT /*tag= e

FT FT /note= "g10 translational enhancer"
 FT RBS 184..188
 FT /*tag= f 197..223
 FT CDS
 FT FT /*tag= g
 FT CDS /note= "mini-clstron"
 FT 227..1651
 FT /*tag= h
 FT /note= "modified doxa gene; the native doxa
 FT coding sequence comprises nucleotides
 FT 383..1651"

XX XX MO9744439-23
 PD 27-NOV-1997.
 XX XX
 XX 22-MAY-1997; 97WO-US08690.
 XX XX
 XX 24-MAY-1996; 96US-0653650.
 XX XX
 PA (OHTS) UNIV OHIO STATE RES FOUND.
 XX XX
 PI Desantel CL, Dickens ML, Strohl WA;
 DR WPI; 1998-018495/02.
 DR P-PSDB; AAM36132.
 XX XX
 PT Producing doxorubicin from daunomycin by daunomycin C14 hydroxylase
 PT - also hydroxylation and oxidation of other anthracycline(s) with
 PT the same enzyme
 PS
 PS
 PS Disclosure; Fig 9; 59pp; English.

CC This DNA sequence comprises a fragment of plasmid PANT199 in
 CC which the Streptomyces sp. strain C5 doxa gene (see AAV01447) is
 CC translationally fused with a leader sequence encoding 6
 CC histidine residues so that the fusion protein can be affinity
 CC purified on a nickel-agarose gel. The doxa gene codes for
 CC daunomycin C-14 hydroxylase (see AAM36128), an enzyme capable of
 CC converting daunomycin to the anticancer agent doxorubicin. Host
 CC cells, especially Streptomyces host cells, transformed with
 CC PANT199 can be used in methods for the production of doxorubicin
 CC from daunomycin or for the hydroxylation and oxidation of other
 CC anthracyclines.
 CC XX

SQ Sequence 2081 BP; 380 A; 688 C; 657 G; 356 T; 0 other;

Query Match 42.6%; Score 1221.6; DB 19; Length 2081;
 Best Local Similarity 94.1%; Pred. No. 9,3e-161;
 Matches 1280; Conservative 0; Mismatches 79; Indels 1; Gaps 1;

Oy 5 ccgacacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 64
 Db 1730 CCGCACCGGGCGACGCGGACCGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1671
 Oy 65 cc-ggtcgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 123
 Db 1670 CCGGGTCGGGTGGCGCCATCAACGAGCAGAGCGGCAAGTGAAGCCCGCTGTCT 1611
 Oy 124 gggccccccttcgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 183
 Db 1610 GGGCCCCCTTCTCGACACGCGCACTCTCTACGCGACGCGCGCGCGCGCGCGCGCGCG 1551
 Oy 184 acctctcgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 243
 Db 1550 ACCTGCTGCGAGTACGCGCATATGTCGCGGATTCAGCTGCGCGGAGACTCTCCCGCA 1491
 Oy 244 tgcagtagtgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 303
 Db 1490 TCGAGTAGTGGCG 1431
 Oy 304 aggcgtcgaggcgctgtagcgagcagcagcagcagcagcagcagcagcagcagcagcagc 363

[illegible]

AAV34001/C	AAV34001 standard; DNA; 1569 BP.
ID	AAV34001
AC	AAV34001;
XX	
DT	25-JAN-1999 (first entry)
XX	
DE	DNA fragment containing S. peuceletus dnru gene.
XX	
KM	dnru; dnrx; recombinant; microorganism; doxorubicin; daunorubicin;
KW	metabolism; baumycin; secretion; anti-tumour compound; ss.
XX	
OS	Streptococcus peuceletus.
XX	
FH	Key
FT	Location/Qualifiers
FT	1..255
FT	/*tag= a
FT	/product= "pdsg gene product"
FT	/transl_except= (pos:34..36, aa:Leu)
FT	252..1115
CDS	/*tag= b
FT	/product= "dnru gene product"
FT	/transl_except= (pos:426..428, aa:Met)
FT	/transl_except= (pos:621..623, aa:Met)
FT	/transl_except= (pos:624..626, aa:Asn)
FT	/transl_except= (pos:750..752, aa:Asn)
FT	/transl_except= (pos:801..803, aa:Asn)
FT	/transl_except= (pos:855..857, aa:Asn)
FT	1219..1569
CDS	/*tag= c
FT	/product= "partial dnru gene product"
FT	/transl_except= (pos:1255..1257, aa:ASP)
FT	/transl_except= (pos:1258..1260, aa:Gly)
FT	/transl_except= (pos:1420..1422, aa:Asn)
FT	/transl_except= (pos:1450..1452, aa:Asn)
FT	/transl_except= (pos:1456..1458, aa:Asn)
FT	/transl_except= (pos:1480..1482, aa:Pro)
FT	/note= "no stop codon is given at the 5' end of the sequence"
FT	
XX	
PD	WO9839458-A1.
XX	
PD	11-SEP-1998.
XX	
PF	05-MAR-1998; 98WO-US03938.
XX	
PR	28-JUL-1997; 97US-0901306.
XX	
PR	06-MAR-1997; 97US-0812412.
XX	
PA	(PHMA) PHARMACIA & UPJOHN SPA.
XX	
PI	Colombo AL, Filippini S, Fonstein L, Hutchninson RC;
PI	Lomovskaya N;
XX	
DR	WPI; 1998-506367/43.
XX	
DR	P-PSDB; AAM68558, AAM68559, AAM68560.
XX	
PT	Production of doxorubicin and daunorubicin - using a recombinant
PT	daunorubicin producing microorganism in which at least one dnru or
PT	dnrx metabolism gene has been inactivated
XX	
PS	Disclosure; Fig 2A-B; 48pp; English.
XX	
CC	This sequence represents an SctI-AatII restriction fragment from the
CC	Streptomyces peuceletus genome, which contains the dnru gene. The
CC	invention relates to a recombinant microorganism which produces
CC	doxorubicin (DOR), in which at least 1 daunorubicin (DAR) metabolism
CC	gene selected from dnru and dnrx, has been inactivated. The microorganism
CC	is preferably S. peuceletus WMH1658, WMH1654 and WMH1662. The function
CC	of the Dar gene is inactivated by insertion of a gene (preferably
CC	neomycin/tetramycin resistance gene aphII) into the daunorubicin
CC	metabolism gene. The blocking of the function of at least one gene
CC	of DAR metabolism can increase DAR and DOR production levels and cause

Query Match: 88.68; Score 2542.2; DB 2; Length 3196;

Best Local Similarity 93.6%; Pred. No. 0;
Matches 2681; Conservative 0; Mismatches 163; Indels 21; Gaps 22

[illegible]

Dh	1825	GGATCGGGCGCAGACGCCGGTGGTCCCTCACACCTGCTCACAGCGGATGAGCTGAACGGACCGCA	1766
Qy	1084	gtcccgagacagggagatgttcagagaccgttcgtccaccccccccgacagcgcgggcgagagat	1143
Dh	1765	gctcccggaacccggagatgtcagacacgctgcacaccccccccgacagcgggcgagat	1706
Qy	1144	cggagtccttaccgaacccggagatcggcgacaccccgcgggcgagggggtatctatgtgtga	1203
Dh	1705	CGGGTCCCTTACGACACCGGGGATCGGCCACGACACCTCGCGGCGAGAGGCGTATATGGTGA	1646
Qy	1204	tgaaccagcgcggtccgcgcgcgcgcgcgcgtctcaaccctcagacacagcgcgccgcctccgga	1263
Dh	1645	TGACCCAGAGCGGGTCCGCCGCCGGGGCGTTCACTCGACGAGAGGGGCCCGCTCTCGGA	1586
Qy	1264	agcgctcgttcacactcgggttcgctcgtatgttcatacttgggaacgcgcgaagtggtcga	1323
Dh	1585	ATGCGTCTGTGACCTCGGGTTTGCGCTGCATGTGTCATATGGGACACGAGAACGGGTGTGA	1526
Qy	1324	cggacaccccgggcggtctgcgcgtcagacagagacccgcgcgcgcgcgcggtatccctc	1383
Dh	1525	CGGCCACCCGGGGGCGCTCCGCCGCTCATGAGCGACCGTCCGCCGCCGGGGTACGCTTC	1466
Qy	1384	ccgcaagttcgaacacacgagaagcccgccccgtcgcggttcagacaggtccgcgcgcgcgc	1443
Dh	1465	CCGAGATTGACACACCGAGAGAGCGCGGCCGCTCGGGGTGACACAGTCCGCCGCCGCC	1406
Qy	1444	cctgggcgtgtcgcgcgcgcgcgtctctcgaagagacccgcgaattcaacggcgcgccagac	1503
Dh	1405	CCTGGGGGTGTGCACAGGGGCTCTGTTCTCGACGAGACCGCCGCAATTGCAGCGGCCCGGAC	1346
Qy	1504	cgtcgcgtcgcagtcgttcgcacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	1563
Dh	1345	CGTGCCTGCGCAGTCTGTGCACGAGAGAACACACAGCGCCCATGTGGGGCCGACCGCCCGT	1286
Qy	1564	gacgcaccagacccctcgggtctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	1623
Dh	1285	CGTGCCTGCGCAGTCTGTGCACGAGAGAACACACAGCGCCCATGTGGGGCCGACCGCCCGT	1226
Qy	1624	gacgcaccttcagtcgcgttcgc	1683
Dh	1225	GACGCGCTTGCGGTCCGCGTGTGCAGCGAGACGCGGAGCGGAGGTCGACAGCGTGAACCCACG	1166
Qy	1684	gcgcgcccgccgatctgcgcgttcggttcaacagctcgcacccagccgcgcgcgcgcgcgcgc	1743
Dh	1165	GGCGGTCCCGCATGTGTCGTGGGTACACAGTTCCACCCACGCGACGACGCGGCGACGCCGT	1106
Qy	1744	cacctccgcgcgcctcagatgactcccttcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	1803
Dh	1105	CACCTCCCGCGCCCTCCATGACGCCCTTGGCGCACACGCGGAACCGCGGCCCGCATGGGTC	1046
Qy	1804	ggcgagaagccgcacatccgc	1863
Dh	1045	GGCGAGACGCGCCATCCGGCGCGAGACCGAGACAGTCCATCGGGGTCAATGATACGTCGCC	986
Qy	1864	gccccgcgccttcgc	1923
Dh	985	ACCGCGCGCTTCGACCCGTTTGGTCACTGCGTGGGCGTGTGGTGGCGCAATACACGCT	926
Qy	1924	ccagaatgc	1983
Dh	925	CCAATATGCGCGCATGCGCGTGTGGTGTGTCGCGGCGCGTACGAGCGGTGTAGGGGT	866
Qy	1984	gtcgatctgtgtgc	2043
Dh	865	GTGATGTGATGGGGGCGACCGCGCGCGACAGGCTTCCGTCGGAGCTGAAGATGATGTA	806
Qy	2044	tcccccgc	2103
Dh	805	TCGCCCGGACCCCGGTCGTGACACAGGTCGCGGTCCACGCGGAAGAGGCCGCTGTAGAA	746
Qy	2104	gtcgcgccgagcgcgacacatcgc	2163
Dh	745	GTGCGCGAGAGCGCGCGACATCGGGCGACCCCGAGAGTTCGAACCTATGCGGGGCGCGCGGC	686

[illegible][illegible]

QY	1744	aacctccggccctccatagatccctcttgcccgagaccggaacgggcccggcgaggtc	1803
Db	1105	CACTCTCCGGCCTTCATATGAGCCCTTGCCCAACACGGGAACGGGGCCCGCATGGGTG	1046
QY	1804	ggcgaagaccgcacatccggccgagaccgaagaaagtcacatcgatgatctacccgc	1863
Db	1045	GGCGAAGACCCCAATCCGGCCGGAAGCCGAGAGAGTCCATCTGGGTGATGATAGCTGCC	986
QY	1864	ggccgcgcctcccgaccgccttgatcagctgctgtgtgtgtgtgtgtgtgtgtgtgtgtgt	1923
Db	985	ACCCGCCCTTCGACCCGTTTGATGATAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	926
QY	1924	ccagatgcccgcacatgcgcgtgctgtgctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt	1983
Db	925	CCAGATGCGCCGCAATGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	866
QY	1984	gtcgaatctgt	2043
Db	865	GTTCGATCTGATGTGAGGGGGGAGACCGGGGGAGACAGGCTTTCCTGTGTGAGCTGTAAAGTATG	806
QY	2044	tccccgcgcgcgcgcgtgtcgtctgaaacaaagctgtgtgtgtgtgtgtgtgtgtgtgtgtgt	2103
Db	805	TCCCCCGGCACCCCGGTCGTGACACACAGGTCGGGTGTCAACGGGAAGAGCCGGTGTACAA	746
QY	2104	gtcgcgcgagggcgcgcacatccgcggcgagaccgaggttgaaacatgcggtggcgcgccg	2163
Db	745	GTCCGCGCCAGCGCGGCGAGATTCGGGCGCAACCCGAGAGTGTGAACATATCGGGGGGGCGCGCGC	686
QY	2164	gaacctgtgtccagatctgtctctcttcgtatgtatctgcgaacagagagctgtgccttcgcga	2223
Db	685	GAACCTGGTCAAGATGTGTTCTTCTTGTGACGAGATCGGCAACACAGCGTGTGCG-----	632
QY	2224	tgaagacggaacatctgcggaatgagacggaacatgtgcggcgggcgccgcgcgcgcgtacgt	2283
Db	633	-----CTTCGCGGATGTGAGACGGGCGATATCGGGCGGGGAGGCGCCCGCGCGGTAGC	586
QY	2284	ggcgaggtgtgc	2343
Db	585	CGCGGCTGTGGCGACCGT	526
QY	2344	cgcgcgtctgcgcgcctcaaccgggaagcagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	2403
Db	525	CGGGGCTCGGGCGCTCATCCGGGGAGACGCGCGCGCGCTGATGATGACGCGCTGGTCA	466
QY	2404	actctctgc	2463
Db	465	ACTCTCTGGCGCGGGCG	406
QY	2464	gcaggaagggggttgaaacccggaagttaagagcagagcgagctgacccgcgtccgaatgcgaatgc	2523
Db	405	GCAAGGAAGGGGTGGAACCGGAATGAGTGAAGCGCAGTGTGCCCGCCCGCATGCGGGGTGC	346
QY	2524	ggacctaacccggggtgatagctgtgacccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	2583
Db	345	GGACCTTCGCGGGGATGTAGCTATCCTGATGAGACATCTCGGGCACCGCGCGCGCGCTCTCG	286
QY	2584	ccgcgcgtcatgatgtgtgcgcgtgtgttgagacgtgcgttacgccttgcgcgcgcgcgcgcgcgc	2643
Db	285	TGGCGGTCAATGATGTTGGCCTCTGTTTGTGAGCTGCCGTACGCCCTGGCGCTGTAGGGGT	226
QY	2644	gaagcgttcgcgcgcgttgagagtcgtccgcgtgtcgaatccgcgcgcgcgcgcgcgcgcgcgcgc	2703
Db	225	GACCGTTCGCGTGTGAATGATGATCCGGGTGCATTCGGGCTCTGTGGTGTACGCGTCCGACAGG	166
QY	2704	tgaagatcacgc	2763
Db	165	TGAAGATCAACCGCCCGCGCGGAGCCGCTCCCGAGACGCCGTGTGCCAAGAGAAAGCCCG	106
QY	2764	cgaggttattgacctgtgatatgtgtgtgcctcgaaccgcgtctctgtgtgtgtgtgtgtgtgtgt	2823
Db	105	CGAATGATGATGACCTGGAAGGGTGGGCTCGAACCCTGCTCTGTGGTGTGTGTGTGTGTGTGTGT	46

|||||
Db 45 ACATCGCCGCGGCGTTCGCCGCGCATGATCATGCGCGGTACC 1

RESULT 3
US-08-396-218-1/c

Sequence 1 Application US/08396218
Patent No. 5695966

GENERAL INFORMATION:

APPLICANT: INVENTI, Augusto

APPLICANT: BREME, Umberto

APPLICANT: COLOMBO, Anna L.

APPLICANT: HUTCHINSON, Charles R

APPLICANT: OTTEN, Sharee

APPLICANT: SCOTTI, Claudio

TITLE OF INVENTION: PROCESS FOR PREPARING DOXORUBICIN

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:
ADDRESS: NIKAI DO, MARCELSTEIN, MURRAY & ORAM

STREET: 655 Fifteenth Street, N. W., Suite 330 - G

CITY: Washington

STATE: DC

COUNTRY: USA

ZIP: 20005-5701

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/396, 218

FILING DATE:

CLASSIFICATION: 536

ATTORNEY/AGENT INFORMATION:

NAME: KITTS, Monica C

REGISTRATION NUMBER: 36,105

REFERENCE/DOCKET NUMBER: P1615-5002

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202/638-5000

TELEFAX: 202/638-4810

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 1269 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

HYPOTHETICAL: NO

ANTI-SENSE: NO

ORIGINAL SOURCE:

ORGANISM: Streptomyces peucetius

FEATURE:

NAME/KEY: CDS

LOCATION: 1..1269

US-08-396-218-1

Query Match 44.2% Score 1269; DB 1; Length 1269;
Best Local Similarity 100.0% Pred. No. 3e-191;
Matches 1269; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

263 gccgaaggtgaacccgcgcacaggaagcgtccgggtggaagcgtcggggcgtcgt 322
|||||
Db 1089 GCCGAAGGTGAACCCGCCGCCACAGAGAGGATCCGGGTGGAAGGCTCGGGGCTGTG 1030
323 atggcgccgtcgt 382
1029 ATGGCGCCGTCGT 970
383 ccggaacgc 442
Db 969 CCGGAACGCCGCCGAGATGATCCTCCGTGACAGCAACCTCCCAACGTTAGGACCGG 910
443 cgggt 502
Db 909 CGGT 850
503 cgcgaagc 562
849 CGCGAGCG 790
563 gtggc 622
Db 789 GTGGCGCGCGAGATGAGCCCGGTATCATGTAGACGACCTGTGTGTGTGTGTGT 730
623 gaactcgc 682
729 GAACCTCGGCTCG 670
683 ggcgtgagtgacagcgttcgaaggaagcgtctccagggcgcgcgcgcgcgcgc 742
669 GCGTAGTGCACAGCGCTTCGAGGAGCAGCGCTTCAGGGCGCGAGGTGTCGCGACCGCC 610
743 ggcgaaggtcgt 802
609 GCGAGGTCCTGCG 550
803 aacgttcgaagc 862
549 AACGCTGACGGGCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 490
863 gatgacacacagc 922
489 GATGACCAACAGCGGAGAGTGTACGCCAGCGCGCGCGCGCGCGCGCGCGCGCG 430
923 ccggc 982
429 CCGGCCGAGGCGGTGCGGAGTTGCGTGAAGCCGCGCGCGCGCGCGCGCGCGCG 370
983 cgtcgc 1042
369 CGTCCGCTCG 310
1043 gtcggccttcgc 1102
309 GTGGCGCTCG 250
1103 ggaacccgtgtccaccccccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1162
249 GAGACCGTGTCCACCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 190
1163 ggaatcgc 1222
189 GGGATTCG 130
1223 cgcgggggcgcgttcacactcgaacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1282
Db 129 CGCGGGGCGGCTTCACCTGACAGACCGGGCGCGCGCGCGCGCGCGCGCGCGCG 70
1283 ctgtgcgtgcatgtgtcatcatatggaacgcgcgcgcgcgcgcgcgcgcgcgcgc 1342
69 CTTGGCGTGCATGATCATCATGAGGACCGCAACGGGTGACGGCGCGCGCGCGCG 10

QY	1343	gccgtcac	1351
Db	9	gccgtcac	1
QY	143	ccgcaactgctgtatcgacagcgagtcgagtcggagcctctgggaaactgtctgacagtaagcc	202
Db	1269	tcagcgacagccagacgagcgagtcgtgtgagccgcgcgcgtctgtggcccccttcgcgacaca	142
QY	83	tcagcgacagccagacgagtcgtgtgagccgcgcgcgtctgtggcccccttcgcgacaca	142
Db	1269	tcagcgacagccagacgagtcgtgtgagccgcgcgcgtctgtggcccccttcgcgacaca	1210
QY	1209	ccgcaactgctgtatcgacagcgagtcgagtcggagcctctgggaaactgtctgacagtaagcc	1150

QY	203	gacatcgttcgcgacatcccaagctcgtggccgagctctccccaatgcagtagtcgcgcgc	262
Db	1149	GATATCTCTGGCGCACTCCAGCTGGGGAGACTCTCTCCGATGCAATGTGGCGCCCTTC	1090
QY	263	gcocaaagtctgcgcgcgcgcacagagagacgtctcggctgtgaaagcgctgcgggcctg	322
Db	1089	GCCGAAGGTGAGCCGCCGCCACAGAGACGCTCCGGGTGGAAAGCGCTGGGGGCTCGTG	1030
QY	323	atgcgcgcctgcagtgctctgtctgcctcgaatgtccacagacacgcgcgcctcgcgcggag	382
Db	1029	ATGGCGGCGCTCGGCTGTTGGAGCCCTCGATGTCACACACACCGGGCGCTCCGCGGCGAG	970
QY	383	ccgcgcgcgcgcgcgcgttgttcaatccctctgtgcgaagcaatcccaacagtgtatgcacccg	442
Db	969	CCGGACCGCCCGCGAATGTGATCTCCGTGGGAGGAACCTCCACACAGTGTAGGAGCACCGG	910
QY	443	cgagtgtagtagcgcgcgcctcctccaaagcgcggagagacgcgcctctgcctgcacacgc	502
Db	909	CGGGTGTAGGCGCAGCCGCTCTCTCCAGAACCGGGAGAGCGCGCTCTCTCGGCATCCGC	850
QY	503	cgcgagcgcgccgcgcgcagagacctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	562
Db	849	CGCGAGCGCGCCCGCCCGAGACCTCCGCGACACAGAACCCAGGAAGAGACCGTGTGTTC	790
QY	563	gtgc	622
Db	789	GTGGCCCGCGCAAGATGAGCCCGGTGATCATGTAGACACTGTGTCTCTCGAATCCGACCTC	730
QY	623	gaactcgcgcgcgcgcgcgcgcctcgtacaaacgcgcgcgcgcgcgcgcgcgcgcgcgcgc	682
Db	729	GAACTCGGCGCTGGCGCGGCTGTATACACACCGCGGCTATGTGGGGTGTCTGTCCGGCG	670
QY	683	ggtctgagtgcaagcctcgc	742
Db	669	GCGCTAGTGCAACCGGCTTCGAGAGACAGGCTCTCAAGGCGCGAGGTGTCGCGCACCGCC	610
QY	743	ggcagaggtccgctgc	802
Db	609	GCGAGGGTCTCTGGCGCTGATCCCCCGCGCTCTCTCGGGACCGCGAGGCTCGAGTCTTAAG	550
QY	803	aacgctgc	862
Db	549	AACGCTGACGGGCTCTGGGGGCGANTCGCGGATGTGGTAGCCGGACACCGAGCATGCTCGCA	490
QY	863	gattgcacaaagc	922
Db	489	GATGACCAACAGCGGGAATGTGTACGGGAAGCCGCGATAGCTCGCGGCTTCCCGA	430
QY	923	ccgc	982
Db	429	CCGGCCGAGAGCGCTCGCGAGATTGGTAGACACCGCCGCGCATGTGCGGCGATGCGATC	370
QY	983	gctccgctcgc	1042
Db	369	CGTCCGCTCGGCGACCGCGCGCGGCTGTGACAGCAGTGTGGATGTCGGCGCAGCGCGCG	310
QY	1043	gtgc	1102
Db	309	GTGGGCTCTGGCGCTGACAGGGAATGACGCTGAAAGGACGACGCTCCGAAACGGGATGTC	250
QY	1103	gagacgc	1162
Db	249	GAGACCGCTGCTCACCCCCCGCCACAGCGCGCGGGGGGCGAAGTCTCGGGTCTTTCACAAACCG	190
QY	1163	ggagatcgc	1222
Db	189	GGGATCGGCGCAGCACTTCGGGGGAGAGGCGTCACTGAGTATGACCCAGCGGGGTCCGC	130
QY	1223	cgcgggggcgcttcaactcgc	1282
Db	129	CGCGGGGCGCTTCACTCTCGAGACCGGGCCCGCTCCCGGAAGCGCTCTGTGACCTCGGG	70


```

Db 410 CGGCCACCCGGCGGCGCTGCGCGTCATGAGGACCCCTC 371
RESULT 7
US-09-096-982-6/c
Sequence 6, Application US/0906982
Patent No. 5962293
GENERAL INFORMATION:
APPLICANT: Strohl, William R.
APPLICANT: Dickens, Michael L.
APPLICANT: Desanti, Charles L.
TITLE OF INVENTION: METHOD OF PRODUCING DOXORUBICIN
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESSES:
ADDRESSEE: CALFEY, HALTER & GRISWOLD
STREET: 800 Superior Avenue, Suite 1400
City: Cleveland
STATE: Ohio
COUNTRY: USA
ZIP: 44114-2688
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/096,982
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Gollick, Mary E.
REGISTRATION NUMBER: 34829
REFERENCE/DOCKET NUMBER: 22727/00131
TELECOMMUNICATION INFORMATION:
TELEPHONE: 216-622-8458
TELEFAX: 216-241-0816
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 3013 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
US-09-096-982-6
RY Match 38.4%; Score 1102.2; DB 2; Length 3013;
t Local Similarity 94.4%; Pred. No. 3.8e-165;
ches 1154; Conservative 0; Mismatches 68; Indels 1; Gaps 1;

```

[illegible]

Search completed: June 11, 2002, 17:59:11
Job time: 3998 sec

Wed Jun 12 08:59:13 2002

us-09-673-254-1.rni

Page 10


```

|||||
17 tmetThrMetGlnArgLysProGluValHisAspAlaPheArgGluVala 34
1251 GCGCGGTCGTCGAGTGTGACGCCCCCGCGGCGAGACCCCGCTGGGTATC 1202
34 lYpValValGluValAsnAlaProAlaGlyGlyProAlaTTPValIle 50
1201 ACCGATGACGCCCTCGCCGCGAGTGTCTGGCCATCCCGGTTCGTAA 1152
51 ThrAspAlaLeuAlaIaArgGluValLeuAlaAspProArgPheVally 67
1151 GGACCCGACCTGCGCCCGCGCTGGCGGGGGTGAGACGCTGCG 1102
67 sAspProAspLeuAlaProAlaIaIaTTPArgGlyValAspAspIlyLeuA 84
1101 ACATCCCGCTCGGAGCTGCGCTCCGTCACGCTCATCGCGCGAGCGC 1052
84 splleProValProGluLeuArgProPheThrLeuIleAlaValaAspGly 100
1051 GAGGCCACGCGCGCTCGCCGCGATCCACGACGACCTGCTCAACCGCG 1002
101 GluAlaHisArgArgLeuArgArgIleHisAlaProAlaPheAsnProAr 117
117 gArgLeuAlaGluArgThrAspArgIleAlaAlaIleAlaGlyArgLeuL 134
1001 CCGGTCGCGGAGCGGACGATGCGATCGCGCGCGATGCGCGCGCTGC 952
117 gArgLeuAlaGluArgThrAspArgIleAlaAlaIleAlaGlyArgLeuL 134
951 TCACCGAACTCGCGGAGCGCTCGCGCGCTGGGCAAAACGCGCGAGCTG 902
134 eutHrgLueLeuAlaAspAlaSerGlyArgSerGlyLysProAlaGluLeu 130
901 ATGCGGCGCTCGCGTACCACTTCGCGCTGTGTGTCATCGAGCTGCT 852
151 lIleGlyGlyPheAlaTyrHisPheProLeuLeuValIleGlyGluLeuLe 167
851 CGGTGCGCGCTCACGATCCGCGGATGCGCGCGAGCGCGCTGACGGTTC 802
167 uGlyValProValThrAspProAlaMetAlaArgGluAlaValaSerValL 184
801 TCAAGCACTCGCGCTCGCGCGCGCGAGACGCGCGGGGTGACGCGACG 752
184 eutYsAlaLeuGlyLeuGlyLysProGlnSerGlyGlyLysAspGlyThr 200
751 GACCTGCGCGGGGCGCTCGCGACCTCGCGCTGAGAGCGCTGCTCT 702
201 AspProAlaGlyGlyValProAspThrSerAlaLeuGlnSerLeuLeuLe 217
701 CGAAGCCGTCACCTACGCGCGGAGACGACCGGACCGACCATGACCGCG 652
217 uGluAlaValaHisSerAlaIaArgArgAsnAspThrProThrMetThrArgY 234
651 TCGCTGACGAGCGCGCGAGCGGAGTGTGCGCTGCTCGACGACGACG 602
234 alLeuTyrGluArgAlaGlnAlaGluPheGlySerValSerAspAspGln 250
601 CTGCTATCATGATCACCGCGGCTCATCTCGCGCGCGACGACGACCG 552
251 lLeuValTyrMetIleThrGlyLeuIlePheAlaGlyHisAspThrThrG 267
551 CTGCTCTCTGGGCTCTGCTGCGGAGGCTCGCGGGCGCGCTGGCGG 502
267 ySerPheLeuGlyPheLeuLeuAlaGluValaLeuAlaGlyArgLeuAlaA 284
501 CGGATCCGACGAGACGCGCTCTCCGGTCTGTCGAGGAGGCGCTGCCG 452
284 lAspAlaAspGluAspAlaValaSerArgPheValGlnGluAlaLeuArg 300
451 TACACCGCGCGGCTCGCTACACGCTTGAGAGTTCGCTGCCACGAGGT 402
301 TyrHisProProValProTyrThrLeuTyrArgPheAlaAlaThrGluVa 317
401 GACCATCGCGGCGCTCGCGCGGAGCGCGGCTGCTGTCGAGAA 352
|||||

```

```

317 lThrIleGlyValaArgLeuProArgGlyAlaProValLeuValaAspI 334
351 TCGAGGCGACCAACACGAGCGCGCGCATCAGACGCGCGCACGCGCTTC 302
334 lGluGlyThrAsnThrAspGlyArgHisHisAspAlaProHisAlaPhe 350
301 CACCGGACCGCTCGCTCGCGGGGCGCTCACCTTGGGCGAGCGGCGGA 252
351 HisProAspArgProSerThrPArgArgLeuThrPheGlyAspIlyProH 367
251 CTACTGCATCGGAGACAGCTCGCCAGCTGAGTCGCGCAGCATGATCG 202
367 sTyrCysIleGlyGlnLeuAlaGlnLeuGlnSerArgThrMetIleG 384
201 GCGTACTGCGACAGAGTTCGCCGAGCGCGACTGCGCGCTGCGTACGAC 152
384 lYValLeuArgSerArgPheProGluAlaArgLeuAlaValProTyrAsp 400
151 GAGTTCGCTGTCGCGGAGGCGCGACGCGCGCTCACGCACT 102
401 GluLeuArgTyrCysArgGlyGlyAlaGlnThrAlaArgLeuThrGluLe 417
101 GCGCGCTCGCTGCGC 86
417 uProValTTPLeuArg 422

```

seq_name: /SIDSL/gcdata/geneseq/geneseq-emb1/AA1998.DAT:AAW36132

seq_documentation_block:

ID AAW36132 standard; Protein: 474 AA.

AAW36132;

21-MAY-1998 (first entry)

N-terminal modified daunomycin C-14 hydroxylase.

Dauinomycin C-14 hydroxylase; doxa gene; doxorubicin; daunomycin;

13-dihydrocarminomycin; carminomycin; anthracycline;

anticancer; cytostatic; cancer; therapy; plasmid pAN199.

OS Chimeric - Streptomycetes sp. strain C5.

XX Chimeric - synthetic.

Key Location/Qualifiers

FT Cleavage-site 31 /note="enterokinase cleavage site"

FT Protein 53..474 /note="native daunomycin C-14 hydroxylase"

PD W09744439-A2.

PD 27-NOV-1997.

PF 22-MAY-1997; 97WO-US08690.

PR 24-MAY-1996; 96US-0653650.

PA (OHIS) UNIV OHIO STATE. RES FOUND.

PI Desanti CL, Dickens ML, Strohl WA;

DR WPI: 1998-018495/02.

DR N-PSDB; AAV01452.

PT Producing doxorubicin from daunomycin by daunomycin C14 hydroxylase

PS - also hydroxylation and oxidation of other anthracycline(s) with

the same enzyme

PS Disclosure; Fig 9; 59pp; English.

CC This protein comprises an N-terminal modified enzyme derived from

the daunomycin C-14 hydroxylase (see AAW36128) of Streptomycetes sp.

CC strain C5. It is encoded by a gene construct (see AAV01452) in
 CC plasmid pAM199. In this plasmid, the doxa gene (see AAV01447) is
 CC translationally fused with a leader sequence encoding 6
 CC histidine residues so that the fusion protein can be affinity
 CC purified on a nickel-agarose gel. Daunomycin C-14 hydroxylase
 CC is a P450-like enzyme capable of converting daunomycin to the
 CC anticancer agent doxorubicin. Host cells, especially Streptomyces
 CC host cells, transformed with pAM199 can be used in methods for
 CC the production of doxorubicin from daunomycin or for the
 CC hydroxylation and oxidation of other anthracyclines.

XX Sequence 474 AA;

Alignment_scores:

Quality: 2093.50 Length: 432
 Ratio: 4.973 Gaps: 1
 Percent Similarity: 97.454 Percent Identity: 93.056

alignment_block:

US-09-673-254-1/rev x AAW36132 ..

Align seg 1/1 to: AAW36132 from: 1 to: 474

1378 GTACCCCGCGCGCGCGCGG...TGCCCTCGTAGAGCGAGCGCCCG 1332
 ||||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 43 ValProTyrGlyAsnSerGluGlyCysLeuMetSerGlyGluAlaProAr 59
 1331 GGTGGCGCTGCAGCCGCTGCGCGCTGCCATGATGACATGACAGCGCAAGC 1282
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 59 gValAlaValAspProPheSerCysProMetMetGlnArgLysP 76
 1281 CGAGAGTCAGAGAGCCCTTCGCGGAGGCGGCGCGCTGCTGAGGTGAC 1232
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 76 ToglValAlaHisSerPalaPheArgGluAlaGlyProValAlaGluValAsn 92
 1231 GCCCGCGCGCGCGCGCGCGCTGCGCTGATCACCAGTACAGCGCCCTGCGCCG 1182
 . 93 AlaProAlaGlyGlyProAlaTPrValIleThrAspAlaLeuAlaAr 109
 1181 CGAGTGTGCGCCGATCCCGCGCTGCTGTAAGAGACCCGACCTGCGCCCG 1132
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 109 gGluValAlaHisAspProArgPheValLysGlyProAspLeuAlaProT 126
 1131 CGCGCTGCGGCGGCGTGCAGCGAGCTGTCAGATCCCGCTCCGAGGTG 1082
 :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 126 hAlaIleTPrArgGlyAlaLysAspGlyLeuAspIleProValProGluLeu 142
 1081 CGTCGCTCAGCGCTCATCGCGCTGAGCGGCGAGGCCACCGCGCGCTGCG 1032
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 143 ArgProPheThrLeuIleAlaValAspGlyLysAspHisArgArgLeuAr 159
 1031 CGCATTCACGACCTGCGCTTCAACCCGCGCGCTGCGCGAGCGGAGCG 982
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 159 gArgIleHisAlaProAlaPheAsnProArgArgLeuAlaGluArgThrA 176
 981 ATCGCATCGCGCGGATCGCGCGCGCTGTCACCGAGTACCGCGAGCGC 932
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 176 sPArgIleAlaIleAlaLysAspArgLeuThrGluLeuAlaAspSer 192
 931 TCCGCGCGTGGCGGCAACCGCGCGAGCTGATCGCGCGCTTCGCTACCA 882
 ||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 193 SerAspArgSerGlyGluProAlaGluLeuIleGlyGlyPheAlaTyrHi 209
 881 CTTCGCGCTGTGGTATCTGCGAGACTGCTGCGTGGCGGTACACCGATC 832
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 209 sPheProLeuLeuValIleCysGluLeuLeuGlyValProValThrAspP 226
 831 CGGCATGCGCGCGAGCGCGTTCACGCTTCACGAGCTGCGCGCTGCGC 782
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 226 rAlaMetAlaArgGluAlaValaGlyValLeuLysAlaLeuGlyLeuGly 242
 781 GCCCGCAGAGCGCGCGGCGTGCAGCGGACGACCTGCGCGGCGGTGCC 732

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 243 GlyProGlnSerAlaGlyGlyAspGlyThrAspProAlaGlyAspValPr 259
 731 GGACACCTCGGCGCGTGGAGAGCTGCTCTCGGAGCCGCTGCATCAGCC 682
 259 OAspThrSerAlaLeuGluSerLeuLeuGluAlaValaHisAlaAla 276
 681 GCGCGACAGCACCCCGCACCATGACCCCGCTGCTATACAGCGCGCGAG 632
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 276 rGArgLysAspThrArgThrMetThrArgValLeuTyrGluArgAlaGln 292
 631 GCCGAGTTCGCGCTGCTCCGAGAGCAGCTGCTATACATGATCAGCGG 582
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 293 AlaGluPheGlySerValSerAspAspGlnLeuValTyrMetIleThrGl 309
 581 GCTCATCTTCGCGCGCGCACGACACCGCGCTCTCTCGGCTTCCTGCG 532
 309 yLeuIlePheAlaGlyHisAspThrThrGlySerPheLeuGlyPheLeu 326
 531 TCGCGAGAGCTCTGCGCGCGCGCTCGCGCGGATGCCGAGAGAGCGCC 482
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 326 eValaGluValaLeuAlaGlyArgLeuAlaAlaAspAlaAspGlyAspAla 342
 481 GTCTCCCGCTTCGTGAGAGAGCGCTCGCTACACACCGCGCGTCCCTA 432
 :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 343 IleSerArgPheValGluGluAlaLeuArgHisThrProValProTyr 359
 431 CACGTTGTGAGAGTTCGCTGCCACGAGAGTGCATCAGCGCGCGCTCGCGC 382
 :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 359 rSerLeuThrArgPheAlaAlaThrGluValaValaIleArgGlyAlaArgL 376
 381 TGCCCGCGGAGCGCGCGCTGCTGCTGATGATGAGGAGCGACACAGCGAC 332
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 376 euProArgGlyAlaProValaLeuValaAspIleGlyGlyThrAsnThrAs 392
 331 GCCCGCATCACAGAGCGCGCGCACGCTTCACCGCGAGCGCTCCCTCG 282
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 393 GlyArgHisHisAspAlaProHisAlaPheHisProAspArgProSerAr 409
 281 GCGCGGCTCACCTTCGCGCGAGCGCGCGCTACTAGCATCGGAGAGCAGC 232
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 409 gArgArgLeuThrPheGlyAspGlyProHisTyrCysIleGlyGluGlnL 426
 231 TCGCCACAGCTGAGTGCAGAGATGATCGCGGTACTGCGAGCAGAGTTC 182
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 426 euAlaGlnLeuGluSerArgThrMetIleGlyValaLeuArgSerArgPhe 442
 181 CCGAGGCGCGACTGCGCGCTGCGCTAGCAGAGATGCGGTGCGCGGAA 132
 :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 443 ProGlnAlaArgLeuAlaValaProTyrGluGluLeuAlaArgTPrCysArgL 459
 131 GCGGCGCGCAGAGCGCGCGCTCACGAGACTGCGCTGCGGTGCGCGC 86
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 459 sGlyAlaGlnThrAlaArgLeuThrAspLeuProValaTPrLeuArg 474

seq_name: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:AAW36128
 seq_documentation_block:
 ID AAW36128 standard; Protein: 422 AA.
 AC AAW36128;
 DT 21-MAY-1998 (first entry)
 XX
 DE Daunomycin C-14 hydroxylase.
 XX
 XX Daunomycin C-14 hydroxylase; doxa gene; doxorubicin; daunomycin;
 KM 13-dihydrocarminomycin; carminomycin; anthracycline;
 KM anticancer; cytostatic; cancer; therapy.
 OS Streptomyces sp. strain C5.
 XX
 PN WO9744439-A2.

XX 27-NOV-1997.
 PD
 XX 22-MAY-1997; 97WO-US08690.
 PF
 XX 24-MAY-1996; 96US-0653650.
 PR
 XX (OHS) UNIV OHIO STATE. RES FOUND.
 PA
 PI Desanti CL, Dickens ML, Strohl WA;
 XX
 XX WPI: 1998-018495/02.
 DR N-PSDB; AAV01447, AAV01451-53.
 XX

PT Producing doxorubicin from daunomycin by daunomycin C14 hydroxylase
 PT - also hydroxylation and oxidation of other anthracycline(s) with
 PT the same enzyme
 PS

PS Claim 9; Page 31-33; 59pp; English.

This protein comprises the daunomycin C-14 hydroxylase (DC14H) of Streptomyces sp. strain C5. DC14H is a p450-like enzyme useful for the production of doxorubicin from daunomycin, and for the hydroxylation and oxidation of other anthracyclines. Its amino acid sequence was deduced from the doxa gene (see AAV01447) identified in the daunomycin biosynthesis gene cluster from C5 (see AAV01453). Plasmid vectors (see AAV01451-52) have been constructed for use in the recombinant production of DC14H in transformed host cells. Claimed methods for production of doxorubicin include: the use of a host microorganism transformed with a plasmid which contains the doxa gene; incubating a culture of Streptomyces sp. strain C5 with daunomycin; and incubating DC14H with daunomycin. DC14H can also be used in claimed methods of producing 13-dihydrodoxorubicin and carminomycin from 13-deoxydoxorubicin, and of producing 13-dihydrodaunomycin and daunomycin from 13-deoxydaunomycin. The methods preclude the use of halogens as required in the current chemical methods.

Sequence 422 AA:

alignment_scores:
 Quality: 2076.00 Length: 422
 Ratio: 5.027 Gaps: 0
 Percent Similarity: 97.867 Percent Identity: 94.076

alignment_block:

US-09-673-254-1/rev x AAW36128 ..

Align seg 1/1 to: AAW36128 from: 1 to: 422

351 GTGAGCGGAGGAGCGCCCGGGGTGGCCGTGACCCGTTGCGTGTCCAT 1302
 ::
 1 MetSerGlyGluAlaProArgValAlaValaAspProPheSerCysProkoe 17
 1301 GATGACCATGACGAGCCCAAGCCGAGGTGACAGACGCTTCCGGAGCGG 1252
 ::
 17 TmetThrMetGlnArgLysProGluValHisAspAlaPheArgGluAlaG 34
 1251 GCCCGGTGTGAGGTGAACGCCCGCGGGGGGAGCCGCGTGGGTGATC 1202
 ::
 34 LyrProValValGluValaHisAlaProAlaGlyGlyProAlaTrpValIle 50
 1201 ACCGATGACGCCCTGCGCCGCGAGGTGTGGCCGATCCCGGTTGCTGTA 1152
 ::
 51 ThrAspAspAlaLeuAlaArgGluValLeuAlaAspProArgPheValIle 67
 1151 GGACCCGACCTCGCCCGCGCGCTGGCGGGGGGTGACAGCGGTCTCG 1102
 ::
 67 SelYProAspLeuAlaProThrAlaTrpArgGlyValaAspAspGlyLeu 84
 1101 ACATCCCGTTCGAGAGTGGCTGCGTTCAGCATCGCGGTGAGCGC 1052
 ::

84 spIleProValProGluLeuArgProPheThrLeuIleAlaValaAspGly 100
 1051 GAGGCCACCGGCGCCTGGCGCGCATCCAGCAGACCTGCTTAACCCGG 1002
 ::
 101 GluAspHisArgArgLeuArgArgIleHisAlaProAlaPheAsnProArg 117
 1001 CCGGCTGGCCGAGGAGCAGATCGCCGCGATCGCCGATCGCCGCGCTTC 952
 ::
 117 GArgLeuAlaGluArgThrAspArgIleAlaIleAlaIleAspArgLeu 134
 951 TCACCGAAGTCCGCCAGCCCTCCGCGCGGTGGGCAACCGCCGAGCTG 902
 ::
 134 eutThrGluLeuAlaAspSerSerAspArgSerGlyGluProAlaGluLeu 150
 901 ATCGGCGGCTTCGCGTACCATTCGCCGTTGGTGTGATCGAGCTGCT 852
 ::
 151 IleGlyGlyPheAlaTrpHisPheProLeuLeuValIleCysGluLeu 167
 851 CGGTGTGCCGCTCACCGATCCGCGGATGGCCCGGAGGCCGTACAGCTTC 802
 ::
 167 uGlyValProValThrAspProAlaMetAlaArgGluAlaValaGlyVal 184
 801 TCAGGCACTCGGCTCGCGCGCGCCGCGAGCGCGGGGGGTACGGCAG 752
 ::
 184 eulYAlaLeuGlyLeuGlyGlyProGlnSerAlaGlyGlyAspGlyThr 200
 751 GACCTGCGCGGGGGGTGGCGGACACCTCGCGGCTGGAGAGCTGCTCCT 702
 ::
 201 AspProAlaGlyAspValProAspThrSerAlaLeuGluSerLeuLeu 217
 701 CGAAGCCGTGCATCAGCCCGCGGAGACGACACCCGACCATAGCCCGG 652
 ::
 217 uGluAlaValHisAlaAlaArgArgLysAspThrArgThrMetThrArgV 234
 651 TGCTTACGAGCGCGCGCGAGCGGAGTTCGCTCGCTCGACAGACAG 602
 ::
 234 alLeuTrpGluArgAlaGlnAlaGluPheGlySerValSerAspAspGln 250
 601 CTGCTGTACATGATCACCGGCGCTCATCTTGGCGCGCCAGCACACCGG 552
 ::
 251 LeuValIleTrpMetIleThrGlyLeuIlePheAlaGlyHisAspThrThrG 267
 551 CTCTTCTGCGGCTTCCTGCTCGCGAGGTCCTGGCGGCGCGCTCGCGG 502
 ::
 267 YSerPheLeuGlyPheLeuLeuAlaGluValLeuAlaGlyArgLeuAla 284
 501 CGGATCGCGAGAGAGACCCGCTCCGCTCGGTTCGTGGAGAGCGCTGGCG 452
 ::
 284 IaAspAlaAspGlyAspAlaIleSerArgPheValGluGluAlaLeuArg 300
 451 TACCAACCGCGGTCCTTACACGTTGTGAGAGTTCGCTGCACAGAGAGT 402
 ::
 301 HisHisProProValaProTyrSerLeuTrpArgPheAlaIleAlaThrGluVa 317
 401 GACCATCGGCGGCTTCGCGCTGCCCCCGGAGCGCGGCTGTGTGAGCA 352
 ::
 317 lValIleArgGlyValaArgLeuProArgGlyAlaProValLeuValaAsp 334
 351 TCGAGGCGACCAACACGAGCGGCGCCATCAAGCGCCCGCCAGCGCTTC 302
 ::
 334 lGluGlyThrAsnThrAspGlyArgHisAspAlaProHisAlaPhe 350
 301 CACCGGAGCCGCTCCCTCGTGGCGGCGGCTCACCTTCGCGAGCGGCCCA 252
 ::
 351 HisProAspArgProSerArgArgArgLeuThrPheGlyAspGlyProHis 367
 251 CTACTGCACTCGGAGAGACGCTCGCCCGACAGTGGAGTGGCAGATGATCG 202
 ::
 367 sTrpCysIleGlyGluGlnLeuAlaGlnLeuGluSerArgThrMetIleG 384
 201 GCGTACTGCGAGACGAGTTCCCGAGGCGCGACATGCGCGCTCGCTACAGAC 152
 ::
 384 lValLeuArgSerArgPheProGlnAlaArgLeuAlaValProTyrGlu 400

151 GAGTTCGGTGTGTCGGAAGGGGCCAGACGGCGGCTCACGCACT 102
|||||
401 GluLeuAArgTrrpCysArgLysGlyAlaGlnThrAlaArgLeuThrAspLe 417
101 GCCCGTCTGGCTGCGC 86
|||||
417 uPrrValTrrpLeuAArg 422

seq_name: /SIDSL/gcdata/geneseq/geneseq-emb1/AA1998.DAT:AAW36134

seq_documentation_block:

ID AAW36134 standard; Protein: 275 AA.

AAW36134:

21-MAY-1998 (first entry)

Streptomyces orfa gene product.

Daunomycin C-14 hydroxylase: doxa gene; doxorubicin; daunomycin;

13-dihydrocarbamimycin; carbimycin; anthracycline;

anticancer; cytostatic; cancer; therapy.

Streptomyces sp.

W09744439-A2.

27-NOV-1997.

22-MAY-1997; 97WO-US08690.

24-MAY-1996; 96US-0653650.

(OHIS) UNIT OHIO STATE. RES FOUND.

Desanti CL, Dickens ML, Strohl WA;

WPI: 1998-018495/02.

N-PSDB; AAV01453.

Producing doxorubicin from daunomycin by daunomycin C14 hydroxylase

f also hydroxylation and oxidation of other anthracycline(s) with

the same enzyme

PS Disclosure: Fig 3; 59pp: English.

This protein is the encoded product of the Streptomyces sp. strain
C5 daunomycin biosynthetic gene cluster orfa (see AAV01453). This
gene lies upstream of the doxa (see also AAV01447) gene. The doxa
gene codes for daunomycin C-14 hydroxylase (see AAW36128), an enzyme
capable of converting daunomycin to the anticancer agent doxorubicin.
Host cells, especially Streptomyces host cells, transformed with
CC plasmids (see AAV01451-52) containing the doxa gene can be used in
CC methods for the production of doxorubicin from daunomycin or for the
hydroxylation and oxidation of other anthracyclines.

Sequence 275 AA:

alignment_scores:

Quality: 1385.00

Ratio: 5.149

Percent Similarity: 97.818

Percent Identity: 94.182

Alignment_block:

US-09-673-254-1/rev x AAW36134

Align seg 1/1 to: AAW36134 from: 1 to: 275

2175 GTGACAGAGTTCGGCGCGCGCGCGCATGTTGACACTCGGTTGGC 2126

1 MetThrArgPheAlaProGlyAlaProAlaTrrpPheAspLeuGlySerPr 17

2125 CGATGTGCGCGCTCGCGCGCTTCTACACCGGCGCTGTTCGGGTGAGCG 2076
|||||
17 oAspValAlaAlaSerAlaAspPheTrrpGlyLeuPheGlyTrrpTha 34
2075 CCACCGTGTGACGAGACCGCGCGCGCGGAGATACAGAGTTTCAGCTCC 2026
|||||
34 lArthrValValSerAspProGlyAlaGlyTrrpGlyTrrpThrPheSerSer 50
2025 GACGGGAGAGTGTGCGCGCGCGGTGCGCGCGCGCGCGCGCGCGCGCTA 1976
|||||
51 AspGlyLysLeuValAlaAlaValAlaArgHisGlnIleAspThrProTy 67
1975 CCACCGCGCGCTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 1926
|||||
67 rHisArgProTyGlyProGlyAsnAspGlnHisGlyMetProAlaIleT 84
1925 GGACCGGTGACTTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 1876
|||||
84 rPthrValTrrpPheAlaThrAsnAspAlaAspAlaLeuThrLysArgVal 100
1875 GAGACCGCT 1826
|||||
101 GluThrAlaGlyLysAspValIleMetThrPrometAspValLeuGlyLe 117
1825 CGCGCGGATGCGCGCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 1776
|||||
117 uGlyArgMetAlaValPheAlaAspProSerGlyAlaAlaPheAlaValT 134
1775 GGCGCAAGGAGTTCATGAGGGCGCGGAGGTGACGGGCGTGGCGCGCTCG 1726
|||||
134 rPArgLysGlyValIleMetGluGlyAlaGluValThrGlyValProGlySer 150
1725 GTGCGGTGGGTGAGCTGTGACCGAGCGCGCGCGCGCGCGCGCGCGCT 1676
|||||
151 ValGlyTrrpValGlyLeuValTrrpAspAspIleGlyThrAlaArgGlyPh 167
1675 CTACCGCGCGCGCGCGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCT 1626
|||||
167 eTyArgAlaThrLeuGlyLeuAlaPheAlaAspThrGlyArgLysGly 184
1625 TCACCGACCGCGGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 1576
|||||
184 AlThrAspProValTrrpHisIleHisAspThrProValAlaGlyThrArg 200
1575 GACGTGGCGTCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTGTTC 1526
|||||
201 GluLeuGlyThrThrGlyAlaValArgProHisTrrpAlaValLeuPheSe 217
1525 CGTGCAGACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTG 1476
|||||
217 ValHisAspCysAspAlaThrValAlaArgAlaValAlaGlyLeuGlyLys 234
1475 CCCTGCAAGACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTG 1426
|||||
234 eValAlaGlnAsnGluProValAspThrProArgGlyArgAlaAspLeu 250
1425 CTGACCGCGCGCGCGCGCGCGCTTCTGCGGTGCGAACTGCGGAGAGGCTA 1376
|||||
251 LeuAspProHisGlyAlaGlyPheSerValValGlyLeuArgGlyAlaTy 267
1375 CCCCGCGCGCGCGCGCGCGCGCTCG 1351
|||||
267 rProAlaAlaAlaAspGlyAlaSer 275

seq_name: /SIDSL/gcdata/geneseq/geneseq-emb1/AA1998.DAT:AAW68559

seq_documentation_block:

ID AAW68559 standard; Protein: 287 AA.

AAW68559:

25-JAN-1999 (first entry)

```

XX S. peuceletius dnrU gene product.
DE
XX dnrU; dnrU: recombinant; microorganism; doxorubicin; daunorubicin;
KM metabolism; baumycin; secretion; anti-tumour compound.
XX
XX Streptococcus peuceletius.
OS
XX
XX Key Location/Qualifiers
FH Misc-difference 59 /note= "encoded by AAC"
FT Misc-difference 124 /note= "encoded by AAT"
FT Misc-difference 125 /note= "encoded by CAC"
FT Misc-difference 167 /note= "encoded by CAC"
FT Misc-difference 184 /note= "encoded by CAC"
FT Misc-difference 202 /note= "encoded by CAC"
FT Misc-difference 202 /note= "encoded by CAC"

W09839458-A1.
11-SEP-1998.
05-MAR-1998; 98WO-US03938.
28-JUL-1997; 97US-0901306.
PR 06-MAR-1997; 97US-0812412.
XX (PHAA ) PHARMACIA & UPJOHN SPA.
XX
XX Colombo AL, Filippini S, Fonstein L, Hutcheson RC;
PI Lomovskaya N;
XX WPI; 1998-506367/43.
DR N-PSDB; AAV34001.
XX
XX Production of doxorubicin and daunorubicin - using a recombinant
PT daunorubicin producing microorganism in which at least one dnrU or
PT dnrU metabolism gene has been inactivated
PS
XX Disclosure; Fig 2A-B; 48pp; English.
XX
XX This sequence represents the dnrU gene product from the microorganism
CC Streptomyces peuceletius. The invention relates to a recombinant
CC microorganism which produces doxorubicin (DOR), in which at least 1
CC daunorubicin (DAR) metabolism gene selected from dnrU and dnrX, has been
CC inactivated. The microorganism is preferably S. peuceletius WMH1658,
CC WMH1654 and WMH1662. The function of the DAR gene is inactivated by
CC insertion of a gene (preferably neomycin/kanamycin resistance gene aphII)
CC into the daunorubicin metabolism gene. The blocking of the function of
CC at least one gene of DAR metabolism can increase DAR and DOR production
CC levels and cause the disappearance of baumycin-like products resulting
CC in DAR and DOR secretion directly into the culture medium. The products
CC can be used in the production of anti-tumour compounds.
XX
XX Sequence 287 AA;
SO

alignment_scores:
Quality: 978.00 Length: 196
Ratio: 5.041 Gaps: 0
Percent Similarity: 98.980 Percent Identity: 97.449

alignment_block:
US-09-673-254-1/rev x AAW68559 ..
Align seg 1/1 to: AAW68559 from: 1 to: 287
2869 CGGTACCCGCGCATGTCATGCGCAGCAACGCGCGCGCATGTTCTG 2820
|||||

```

```

92 ArgTyrProArgIleAspValMetAlaSerAsnAlaGlyGlyMetPheTr 108
2819 GTGCGCACACAGACAGGACGGCTTCAGGGCCACCATTCAGTCAATC 2770
|||||
108 pSerArgThrThrGlnAspGlyPheGlnAlaThrIleGlnValMetA 125
2769 ACCTCCAGGCTTCCTGCTGGCACGGCTGCGGGAGCGGCTCGCGGGC 2720
:::|||||
125 snLeuAlaGlyPheLeuLeuAlaArgLeuLeuArgGlnValGleAlaGly 141
2719 GGGCGGCTGATCCTCACTCGTCGACGCGTACACCGAGGCGCGATCGA 2670
|||||
142 GlyArgLeuLeuLeuThrSerSerAspAlaTyrThrGlnIleArgIleAs 158
2669 CCGGAGACAGCTCAACGCGGACCGCTCACCGCTACAGCGCGCGGACGGCT 2620
|||||
158 pProAspAspLeuAsnGlyAspArgAsnAlaGlyTyrSerAlaGlyGlnAlaTr 175
2619 ACGGACAGTCCAAACAGACGACATCATGACGCGGCGGAGCGCGCAGG 2570
|||||
175 TyrIleThrSerIysGlnAlaAsnIleAsnThrAlaAlaGlnAlaIleArg 191
2569 CGCTGGCCGAGACGTGCTGGCGGTACACTATCACCCGCGTGAGGTCCGAC 2520
|||||
192 ArgTyrProAspValLeuAlaValSerTyrAsnProGlyGlnValArgTh 208
2519 CCGCATTCGACGCGGACGCGTCCGCTGCTCTACTTCCGGTTCAACCCCT 2470
|||||
208 rArgIleGlyArgGlyThrValAlaSerSerTyrPheAspPheAsnPro 225
2469 TCCTGCGCTCCGCGGAAAGGCGCGGACACCCCTGTCGTGGCTCGC 2420
|||||
225 heLeuArgSerAlaAlaIleAlaIleAlaAspThrLeuValTrpLeuAlaSer 241
2419 GCGCGCGCGGAGAGTTGACACGCGGCGCTACTACAGCGACCGCGGCT 2370
|||||
242 AlaProAlaGlnGlnLeuThrThrGlyGlyTyrTyrSerAspArgArgLe 238
2369 GTCCCGGTGAGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCT 2320
|||||
258 uSerProValSerGlyProThrAlaAspAlaGlyLeuAlaIleAlaLysLeuTr 275
2319 GGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2282
|||||
275 rpgIuAlaGlyAlaAlaAlaValAlaGlyAspThrAlaHis 287

seq_name: /SIDSL/gcgcdata/geneseq/geneseqp-emb1/AA1998.DAT.AAW36133
seq_documentation_block:
ID AAW36133 standard; Protein; 193 AA.
XX
XX AAW36133;
XX
XX 21-MAY-1998 (first entry)
DT
XX
XX Streptomyces orfi gene product.
DE
XX
XX Daunomycin C-14 hydroxylase; doxa gene; doxorubicin; daunomycin;
KM 13-dihydrocarbamimycin; carminomycin; anthracycline;
KM anticancer; cytostatic; cancer; therapy.
XX
XX Streptomyces sp.
OS
XX
XX W09744439-A2.
PD
XX
XX 27-NOV-1997.
XX
XX 22-MAY-1997; 97WO-US08690.
XX
XX 24-MAY-1996; 96US-0653650.
XX
XX (OHIS ) UNIV OHIO STATE. RES FOUND.
XX

```

PI Desanti CL, Dickens ML, Strohl WA;
 XX
 DR WPI: 1998-018495/02.
 DR N-PSDB: AAV01453.
 XX
 XX Producing doxorubicin from daunomycin by daunomycin C14 hydroxylase
 PT - also hydroxylation and oxidation of other anthracycline(s) with
 PT the same enzyme
 XX
 PS Disclosure: Fig 3; 59pp; English.

XX This protein is the encoded product of the Streptomyces sp. strain
 CC C5 daunomycin biosynthetic gene cluster orf1 (see AAV01453). This
 CC gene lies upstream of orfA and doxa (see also AAV01447). The doxa
 CC gene codes for daunomycin C-14 hydroxylase (see AAV36128), an enzyme
 CC capable of converting daunomycin to the anticancer agent doxorubicin.
 CC Host cells, especially Streptomyces host cells, transformed with
 CC plasmids (see AAV01451-52) containing the doxa gene can be used in
 CC methods for the production of doxorubicin from daunomycin or for the
 CC hydroxylation and oxidation of other anthracyclines.

SQ Sequence 193 AA;

alignment_scores: length: 192
 Quality: 961.00 Gaps: 0
 Ratio: 5.005 Percent Identity: 96.354
 Percent Similarity: 100.000

alignment_block:
 US-09-673-254-1/rev x AAV36133

Align seg 1/1 to: AAV36133 from: 1 to: 193

2860 CGCATGATGTCATGGCCAGCAACGCCGGCGGCGATGTCGTGCGGCAC 2811
 1 Argilepsvalmetlaaglyasnalaaglymetpetherargrh 17
 2810 CAGGACCCAGGAGGGTTCAGAGCCCATCCAGTCAATCAGCTCGCAG 2761
 17 rtrththglnaspolyphneglualatrrleuglnvalasnhislsleuadag 34
 2760 GCTTCCTGTCGCGACGCGCTGCGGAGCGGCTCGCGGCGGCGGCGCTG 2711
 34 lypheleuleuallatargyleuallatargyleuallatargyleu 50
 2710 ATCTCTACCTCTGTCGAGCGGTCACCCAGCGCGGATCGACCCGGACGA 2661
 51 lleleuthrseraseraspalaatytrthglnglyargileasproaspas 67
 2660 CCTCAAGGCGGACCGCTACCGGCTACAGCGGCGGCGGCGGCGGCGGCG 2611
 67 pleuasngelyasparghlsarqytrserlaaglyglnalatyglythr 84
 2610 CCAACAGCGCCACATCATGACCGCGGCGGAGCGCGGCGGCGGCGGCG 2561
 84 erlysglnalalasnllmethrhalatrrglnalalalaaargatrrpro 100
 2560 GACGTGTCGGCGGTACCTATACCCCGGCGGATAGTCCGACCCCGCATCG 2511
 101 aspvallleuthrvalsertryhlsproglylvalalagtrhrrarglle 117
 2510 ACGGGGACGGGCGCTGCTCTACTTCGGGTCAACCCCTTCCTGCGGCT 2461
 117 yargelythrvalalaserthrtyrpharphasnprophelenuar 134
 2460 CCGCGCGCAAGGCGCGCAGACCCCTGCTGTCGCGGCTCGCGGCGGCG 2411
 134 eralalalysglyalasprrleuvaltrpleuallalalalalprola 150
 2410 GAGGAGTTGACACGCGGCGGCTACACGCGGCGGCGGCGGCGGCGGT 2361
 151 glngluleuthrthrglyglytyrtyrserasparatargleuserprova 167

2360 GAGGCGCCGACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2311
 167 lserglyprrthralaspalaaglyleuallatalslsleutrrglnalal 184
 2310 GCGCGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 2285
 184 eralalalalvalglnyasprrthrala 192

seq_name: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:AAW68560

seq_documentation_block:

ID AAW68560 standard; Protein: 117 AA.

AC AAW68560;

DT 25-JUN-1999 (first entry)

DE S. peuceitius partial dnrv gene product.

XX dnrv: dnrv; recombinant; microorganism; doxorubicin; daunorubicin;
 XX metabolism; daunomycin; secretion; anti-tumour compound.

XX Streptococcus peuceitius.

OS Key Location/Qualifiers

FT MISC-difference 13 /note= "encoded by GAG"

FT MISC-difference 14 /note= "encoded by CTC"

FT MISC-difference 68 /note= "encoded by CAC"

FT MISC-difference 78 /note= "encoded by CAC"

FT MISC-difference 80 /note= "encoded by ATG"

FT MISC-difference 88 /note= "encoded by TTC"

XX W09839458-A1.

XX 11-SEP-1998.

XX 05-MAR-1998; 98WO-US03938.

XX 28-JUL-1997; 97US-0901306.

XX 06-MAR-1997; 97US-0812412.

XX (PHAA) PHARMACIA & UPJOHN SPA.

XX Colombo AL, Filippi S, Feinstein L, Hutchinson RC;

XX Lomovskaya N;

XX WPI: 1998-506367/43.

XX N-PSDB: AAV344001.

XX Production of doxorubicin and daunorubicin - using a recombinant

XX daunorubicin producing microorganism in which at least one dnrv or

XX dnrv metabolism gene has been inactivated

XX Disclosure: Fig 2A-B; 48pp; English.

XX This sequence represents a partial dnrv gene product from the
 CC microorganism Streptomyces peuceitius. The invention relates to a
 CC recombinant microorganism which produces doxorubicin (DOR), in which
 CC at least 1 daunorubicin (DNR) metabolism gene selected from dnrv and
 CC dnrv, has been inactivated. The microorganism is preferably S. peuceitius
 CC WMH1658, WMH1654 and WMH1662. The function of the DNR gene is inactivated
 CC by insertion of a gene (preferably neomycin/Kanamycin resistance gene
 CC aphII) into the daunorubicin metabolism gene. The blocking of the
 CC function of at least one gene of DNR metabolism can increase DNR and DOR
 CC production levels and cause the disappearance of daunomycin-like products
 CC resulting in DNR and DOR secretion directly into the culture medium.


```

648 gAtgAtgAlaAtgAtgGlnHisAspArgGlyGlyLeuAspArgGly 665
962 ..CGGCGCGTGTCTCCAGCACTCCGAGCGCTCCGGCGGTCGGCGA 915
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
665 iAtgProSerAtgGlnAlaAtgArgGlnAlaLeuGlyLeu 681
914 ACCGGC.....CGAGCTATGGCGGCTTCGGTACCACT 880
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
682 ProGlyGlySerGlnGlyHisAlaValArgHisArgLeuProLe 698
879 TCCGCTGTGTGTCATCGAGCT..... 855
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
698 uAtgProProLeuHisArgGlyAlaGlyAlaLeuArgProGlyG 715
854 ..GCTCGGTGTCCGGT...CACGATCCGGGATGGCCCGGAGCGCT 810
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
715 LyAlaAtgAlaAspGlyProArgArgProGlyGlnProGlyGly 731
810 ..... 810
732 LeuGlyProGlyProAlaArgGlyProHisGlyArgAspGlyArg 748
809 .....CAGCTTCTCAAGGCACTCGGCTCGGC 782
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
748 gLeuAtgGlnValHisAlaAspGlnArgAspProLeuGlnGlyPro 765
781 GCGCGCGAGAGCGCGGCGTGCAGCGACCGCTCCGGCGGCGCTGC 732
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
765 InProAlaGlnArgArgGlnAlaArgAlaArgProAlaProValArg 781
731 GACACACTCGGCGCTGAGAGCTGCTCTCGAAC..... 696
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
782 GlyThrArgAlaProGlyGlnGlyProArgGlyProGlyProHis 798
695 .....CGTGCACTCAGCCCGCGGCAACAGACACCC..... 666
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
798 gAlaHisAlaThrLeuGlnProArgHisLeuHisGlyArgValArg 815
666 ..... 666
815 sProGlnAlaLeuArgGlyAspProArgGlyGlnGlyProGlyLeu 831
665 .....GACCATGACCGCGCTGTCAGAGCGCGCGC 634
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
832 AlaGlyProValLeuLeuGlnHisGlnGlyArgAlaLeuArgGly 848
633 AGCGCGA..... 627
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
848 gGlyArgArgHisAlaGlnAspArgAspGlyLeuProAlaGlyArg 865
626 .....GTTCGGCTCGCTCCGAGCAGCAGCTGTACATGATCAGCG 582
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
865 rGlnAlaValArgGlyValProArgGlyProValGlnProGlyAsp 881
581 GCTCATCTT.....CGCGCGCAGCAGAC..... 558
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
882 GlnHisLeuGlnGlyGlnGlnHisArgArgGlyProArgHisAla 898
557 .....CAGCGCTCTCTCTCG 542
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
898 gGlyGlyArgGlyLeuLeuGlnArgValHisProHisLeuAlaVal 915
541 GCGTCCCTGTCGCGAGGCTCCGGCGCGCTCCGGCGGAGCGCGA 492
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
915 LHisAlaArgArgArgSerGlyLeuArgProSerGlyProAlaArg 931
491 CGA.....GACGCGCTCTCCGGTCTGTGAGAGCGCTGCGCT 451
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
932 HisHisAlaLeuGlyArgGlyProAlaArgGlnAlaGlyAlaArg 948
450 ACCACCGCGC.....GCTGCCCTACAGTGTGAGGCTGCGTCC...C 410
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
948 aProGlnAlaLeuGlnArgProHisHisLeuArgAlaGlyArgAla 965

```

```

409 ACGGAGTGCACCATTCGGCGCTCCGCTGCCCCCGGAGCGCC..... 366
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
965 iSclValAlaLeuArgArgHisProGlnAlaProAlaArgAlaSerVal 981
365 .....GCTGCTGTGACATCGAGGCGACACACCGGAGCGCGCTATC 322
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
982 ProArgGlyGlnGlyGlnHisGlyAlaHis...HisArgAlaGlnPro 997
321 ACGAGCGCGCGCGCTTCACCGCGGACCGCTCCGTCGGCGGCGTC 272
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
997 gArgAspGlnGlyArgGlyProArgAspArgProArgProGlyGly 1013
271 ACCCTTCGGCGAGCGCGCGCTACTGTCATCGGGGAGCACTCCGCGAG 222
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1013 rGLeuArgArgArgHis.....AspArgGlnHisGlyHisAlaGly 1026
221 GAGTCCGCGACAGATATCGCGCTACTGCGGACAGAGTTCCCGAGCG 172
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1027 Gly.....GlyArgAlaArgArgArgGlyProHisGlyPr 1038
171 GACTGCGCGTCCGTCACGACGAGCTTCGCGTGGCGG..... 135
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1038 oValProArgGlyAlaAlaArgValAlaThrProProProThrGlnAsp 1055
134 .....GAAGGCGCGCGAGACGCGCGCGCTCACCGAACTCCGCTGCT 90
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1055 LuAspGlnGlyLeuPro..... 1060
89 GCGTATATCCGCGCGCGCGAGCGGACGCGCGCGCGCGCGCG...AC 43
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1061 .....TrpArgArgArgArgProCysAlaGlyAlaProProArgCys 1075
42 CGCGGCGTGGCGCGCTCCGCTGCTACCGCGTGGC 6
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1075 sGlySerProAlaValProArgSerSerAspGlyAla 1087
seq_name: p1r2:J00405
seq_documentation_block:
hypothetical 119.5k protein (uvra region) - Micrococcus luteus
N:Alternate names: ORF 1 protein
C:Species: Micrococcus luteus, Micrococcus lysodeikticus
C>Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 04-Feb-2000
C:Accession: J00405
R:Shiota, S.; Nakayama, H.
Mol. Gen. Genet. 217, 332-340, 1989
A:Title: Micrococcus luteus homolog of the Escherichia coli uvra gene: Identification
A:Reference number: S04781; MUID:89364717
A:Accession: J00405
A:Molecule type: DNA
A:Residues: 1-1106 <SH1>
A:Cross-references: EMBL:X15867
A>Note: this reading frame extends between two stop codons and does not begin with a
C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homolo
alignment_scores:
Quality: 560.00 Length: 1168
Ratio: 1.131 Gaps: 78
Percent Similarity: 42.380 Percent Identity: 30.908
alignment_block:
US-09-673-254-1 x J00405 ..
Align seg 1/1 to: J00405 from: 1. to: 1106
10 CGGGTACACGCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 59
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
106 ProAlaGlyArgHisGlyArgValHisGlySerValArgLeuGlyGln 122
60 CGGTCCCGGT.....CGCGC.....CGGC 79

```

```
122 IleProGlyLeuArgHisAspLeuArgArgGlyProAlaAlaLeuArg 139
      |||
      |||
80 GGATGAGCG.....CAAGCAGAGCGGACATT 105
      |||
139 rGValAlaLeuLeuLeuArgProHisValProGlyProGlyGlnAl 155
      |||
106 CGGTAGAGCGCGCGTCTGGGCGCCCTTCGGCAGCAGCGCAATCGTGG 155
      |||
155 aGlyArgGlyLeuHisArgGlyProValProGlyArg.....V 168
      |||
156 TAGGCGAGCGCGCATCGCGCGTGGGGAACGTCTCGCAGTACGCCGAT 205
      |||
168 aHisArgProGlyValHisGlnProGlnProAlaLeuHisSerGlyAsp 184
      |||
206 CATCGTGGCGGACCTCAGCTG.....CG 228
      |||
185 HisHisArgAspLeuArgLeuHisAlaProAlaLeuGlyThrCysArgI 201
      |||
229 CGAGCTGCTCCCGATGCAGTAGTCGGCGCGCTCGCGGAGGTGAGCGG 277
      |||
201 yAlaAlaLeuPro...AlaValArgArg.....AlaGlyGlnProAl 214
      |||
278 .....CGGCGCAGAGGAGCGGTCCGGGTGAA 304
      |||
214 laAspProAlaAlaAspArgGlyProAlaArgGlyValAlaAlaArgAlaHis 230
      |||
305 .....GGCGTGGC 312
      |||
231 ProLeuProGlyAlaArgAlaArgGlyProArgProGlnGlyArgValAr 247
      |||
313 GGGCGTGTGATGCGCGCGCGTGGTGGT..... 343
      |||
247 gGlyProValGlnGlyProValHisAlaGlyLeuArgArgArgGlyArgG 264
      |||
344 .....GCCCTCGATGTC 355
      |||
264 lyAspArgProAlaLeuGlyProAlaArgAlaGlnGlnAlaGlyGlnAla 280
      |||
356 CAGCAGACCGCGCGCGTCCGGCGAGCGGAGCGCGCGATGTCACCT 405
      |||
281 His...HisArgArgArgGlyProProArgHisGlnGlyGlnHis... 295
      |||
406 CGGTGGCAGCGAAGCTCCAGCAAGCTGTAGGGCAGCGCGGTGTAGCGC 455
      |||
296 .....ProProAlaProHisArgLeuGlyLyAspArg..ProGlnAl 309
      |||
456 AGCGCTCTCTCAGACAGCGGAG..... 480
      |||
309 aGlyLyArgProArgArgGlyArgValArgLyArgGlyAlaGlyArgG 326
      |||
481 .....GGCGTCTCTCGCGCATCCGCGCGGAGCG 510
      |||
326 rGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 342
      |||
511 GGGCGCGCAGGACCT.....CGCGGAGCAGAGCGCGAGAGAGAG 551
      |||
343 GlnProProValProLeuValLeuArgGlnAlaLeuLeuProGlnArgAl 359
      |||
552 CGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 582
      |||
359 aArgAlaAspArgGlyArgAspArgAlaAlaLeuValLeuLeuGlnGlnP 376
      |||
583 .....CGGTGATCATGTAGACGAGCTGAT 606
      |||
376 roValArgArgValProArgValHisArgHisArgLeuProProAlaGly 392
      |||
607 .....CTCGGAGACCGAGCGCACTCGCGCTCGCGCGGCGGCTGG 644
      |||
393 GlyProGlyProArgArgGlnArgArgAlaValProAlaArgGlyArg 409
      |||
645 T.....ACAGCAGCGGCGTCACTGTCGGGCTGTCTGTCGCGCG 682
      |||
```

```
409 gArgArgAlaValValAlaArgGlnValHisLeuGlyLeu..... 422
      |||
683 GGCTAGATGACAGGCTTCGAGAGCAGAGCTTCAGAGCGCGGAGTGTCCG 732
      |||
423 ..LeuAlaAlaArgAlaArgArgAlaGlnGlnGlnGlnGlnGlnGln 438
      |||
733 GCAGCGCGCGCGGAGGATCGG.....TGGCTTCACCGCG 767
      |||
439 GlnHisProValGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 455
      |||
768 CGGTCTGTGCGGCGCGCGAGCGGAGTGTGTGTGTGTGTGTGTGTGTGT 817
      |||
455 aArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 472
      |||
818 GCGGCGCATCGCGGATCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 867
      |||
472 rGlnAla.....ProLeuHisHis..... 477
      |||
868 CCAACAGCGGGAAGTGTACGGCAAGCGCGCATGCTCGCGCGGTTTG 917
      |||
478 .....GlyLeuArgGlyArgHisProLeuArgAspAlaGln 489
      |||
918 CCCAGCGCGCGAGGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 967
      |||
489 nAlaArgGlyAspArg.....ValGlyArgArgP 499
      |||
968 CGCGCGATCGCATCCGTCGCGCGCGCGCGCGCGCGGTGTGTGTGTGT 1017
      |||
499 roArg.....AlaLeuArgValValHisAlaGly...AspPro 510
      |||
1018 GTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1067
      |||
511 ValProGlyValProArgGlyProProGlnProHisGlyProGlnArg... 526
      |||
1068 AGCGTGAACGGACGACCTCCGGAACGGGATGTGTGTGTGTGTGTGTGT 1117
      |||
527 .....AlaArgGlyArgProValHisArgGlyArgHisP 538
      |||
1118 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1143
      |||
538 roProAlaHisAlaArgGlnHisGlyValLeuLeuGlnValAlaAlaAsp 554
      |||
1144 .....CGGGTCTCTCAGCAAGCGGATCGCGCAGC 1175
      |||
555 GlyProGlyAlaAlaAspArgGlyProGlyAlaGlnGlyAsp...ProG 570
      |||
1176 ACCTGCGGG.....CGAGGCTCATGTGTGTGTGTGTGTGTGTGTGT 1219
      |||
570 yProAlaGlyValProAlaGlyArgArg.....ProArgValProG 584
      |||
1220 GCGCGCGG.....GGCGTTCACCT..CGAGCAGCGGCGCGG..... 1254
      |||
584 InProArgAlaAlaGlyArgHisProLeuArgArgGlyProAlaHis 600
      |||
1255 CTTCCCGGAAGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1304
      |||
601 ProProGlnHisThrAspArgLeuArgAla..... 610
      |||
1305 GGAACAGCGAAGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1354
      |||
611 .....GlyArgArgProLeu...ArgProArgArgAla..Va 621
      |||
1355 GCAGCGCGCGCGCGCGGAGTACCCCTCCCGCAGTTCAGCAGCAGAGA 1404
      |||
621 HisArgProAlaProAlaGlnGlnProProHisArgAspPro.... 636
      |||
1405 GCGGCGCGCGCGCGGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1454
      |||
637 .....ProAlaProAlaGlyProArgGlnHisProHisArg... 648
      |||
1455 CGCGCGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1498
      |||
649 ArgArgAlaArgArgGlyHisAspArgArgGlyGlyLeuAspArgGlyHis 665
      |||
```


alignment_scores:

Quality: 538.00 Length: 1034
 Ratio: 1.231 Gaps: 63
 Percent Similarity: 42.263 Percent Identity: 30.077

alignment_block:

US-09-673-254-1/rev x S27923 ..

Align seg 1/1 to: S27923 from: 1 to: 924

```

2804 CAGAGAGGGTTGAGGCCACCATCCAGTCAATCACTTCGACGGCTTCC 2755
      |||||
12 ProGlySerGlyLeuGlyAlaHisPro.....HisProArgArg..... 24
2754 TCGTGGACGGCTGCTGCGGAGCGGCTCGGGGGGGGGGGCGGATTC 2705
      ::|||::|||
25 .SerGlyAlaAlaAspProAlaAspProValGlyHisProAlaAlaProA 41
2704 ACCTGCTCGACGCGTACACCGGCGCGAT.....CGA 2670
      ::|||::|||
41 rgaAa.....ProGlyProGluProArgThrArgLeuGln 52
2669 CCGGAGACGACCTACACG.....CGACGTCACCGCTACAGCGCGGCC 2626
      |||||
53 ProAlaThrProArgArgSerGlyAlaAlaAspProAlaAsp..ProVal 68
2625 AGGGGTACGGAGCGTCCAAACAGGCGCAACATCATGACCGCGCGAGGCC 2576
      ::|||
69 GlyHisProAlaAlaProAlaGlyAlaProGlyProGluProArgThrArg 85
2575 GCCAGGGCGTGGCGGACGTCGTGGCGGTACATACCCCGGTGAGGT 2526
      |||||
85 uGlnProAla.....ThrProArgArg 93
2525 CCGGACCGCGATCGGAGCGGCGACGCTCGCTCTCTACTCCGGTCA 2476
      |||||
93 er.....GlyAlaAlaAspProAlaAspProValGlyHis 104
2475 ACCGCTTCGCTCGCGCGGAGGCGCGACACCTCGTGTGCTG 2426
      ::|||
105 ..ProAlaAlaProArgAlaProGlyProGluProArgThrArgLe 119
2425 GCGTCCGCGCGCGCGAG.....AGTTGACACGCG..... 2395
      ::|||
119 uGlnProAlaThrProArgArgSerGlyAlaAlaAspProAlaAspProV 136
2394 .....GGGCTACTACAGCGACCGCGGCTGTCCCGGTGAGCG 2356
      |||||
136 aAGlyHisProAlaAlaProArgAlaProGlyProGluProArgThrArg 152
2355 GCCGACCGCGCGCGCGCTCGCGGCGAAGCTCT..... 2320
      ::|||
153 LeuGlnProAlaThrPro.....ArgArgSerGlyAlaAlaAspProAl 167
2319 .....GGGAGCGGCGCGCGCGCTCG..... 2296
      |||||
167 aAspProValGlyHisProAlaAlaProArgAlaProGlyProGluProA 184
2295 .....GCGACACCGCGCTGACGGCGGCGCGCGCGCC 2263
      |||||
184 rGThrArgLeuGlnProAlaThrProArgArgSerGly..AlaAlaAspPr 200
2262 CGCCCGCATGTCGCTCATCGCGAGATGTCGCTCATCCGCGAGCG 2213
      |||||
200 oAla.....AspProValGlyHisProAlaAlaProAlaProArgAla. 212
2212 AGACGCTGTGTGCGGATTCATGAAGAAGACGATTCGTGACACGAGTTCG 2163
      |||||
213 .....ProGlyPro 215
2162 CGCCCGGCGCGCGCGANGTTCGACCTCGGTTCCCGCGATGTGCGCGCC 2113

```

```

216 GlnPro.....ArgThrArgLeuGlnProAlaThrProAr 227
      ::|||
2112 TCGGCGG...ACTTCTACACCGGCTGTTCGGCTGACCGCCACCGTGGT 2066
      |||||
227 gArgSerGlyAlaAlaAspProAlaAspProValGlyHisPro.....A 242
2065 CAGGACCGCGCGCGCGGCGATACAGACGTTGACGTCGCGAGGAAAC 2016
      ::|||
242 IaAlaProArgAlaProGlyPro..... 249
2015 TCGTGGCGGCTGCGCGCGCGACGATGACACCCCTACCGCGCGG 1966
      ::|||
250 .....GlnProArgThrArgLeuGlnProAlaThrProArgAr 262
1965 TACGGCGCGCGGAGACACGACGACGCGCATGC...CGGCATCTGGACCGT 1919
      ::|||
262 gSerGlyAlaAlaAspProAlaAspProValGlyHisProAlaAlaProA 279
1918 GTACTTGGCACCGGACGAGCGCGGACGCTGACCAAGCGGGGTGAGACG 1869
      ::|||
279 rGAlaProGlyProGluProArgThrArgLeuGlnProAlaThrProArg 295
1868 CCGGCGCGCGAGGTCAATGACTCCGATGACGTCCTCGGCTCGCGCG 1819
      |||||
296 ArgSerGlyAlaAlaAsp.....ProAlaAspProValGly 307
1818 ATGGCGCTTTCGCGGACCGCGCGCGCGCGCGCGCTCGCGCGGCA 1769
      ::|||
307 yHisProAlaAlaProAlaProGlyProGluProArgThrArgLeuG 324
1768 GGGAGTCATGAGAGCGCGGAGGTGACGCGCTCGCGCTCGCTGCGCT 1719
      ::|||
324 InProAlaThrProArgArgSerGlyAlaAlaAspProAlaAspProVal 340
1718 GGGTCAAGCTGTGACCGACGACATGGGCGCGCGCGGACTTGTACCG 1669
      |||||
341 GlyHisProAlaAlaProArgAlaProGlyPro..... 351
1668 GCGACCGCTCGGCTCGCGCGCGCGACGCGGATGAAAGCGGTACCGA 1619
      ::|||
352 .GlnProArgThrArgLeuGlnPro.....AlaThrProA 363
1618 CCGGCTGTGACACATGC.....GTGACACACCGGTGC... 1587
      |||||
363 rGArgSerGlyAlaAlaAspProAlaAspProValGlyHisProAlaAla 379
1586 .....CCGGACCGCGAGACTGCGCGCTCGCGCGCGGCTA 1552
380 ProArgAlaProGlyProGluProArgThrArgLeuGlnProAla..... 394
1551 CCGCGCGACGTGGCGCGCTGCTCGCGCGTGCACGACGTGCGAGCGACG 1502
      ::|||
395 .....ThrProArgArg 399
1501 CCGGCGCGCGCTGAACTCGCGCGCTCCGTCGAGAACGAGCCCGGAA 1452
      |||||
399 erGlyAlaAlaAspProAlaAspProValGlyHisProAlaAlaProArg 415
1451 CCGCGAGGGGGCGCGGCGGACGCTGTCGACCGCGAGCGGCGCGCTTC 1402
      |||||
416 AlaProGlyProGluProArg.....ThrArgLeuGlnProAla... 428
1401 TCGGTGTGCACTGCGGGAAGGGGTACCCCGCGCGCGGCGGTCTTC 1352
      |||||
429 .....ThrProArg..... 431
1351 GTGAGCGCGAGCGCGCGCGGTGCGCGCTGACCGCTGCGCTGCCAT 1302
      |||||
432 .....ArgSerGlyAlaAlaAspProAlaAspProVal..... 442
1301 GATGACCATGACGCGCAAGCCGAGGTGACGAGCGCTTCGCGAGGCGG 1252
      |||||

```

```

443 ...GlyHisProAla.....AlaProAlaProGly.... 452
1251 GCCCGGTCTCGAGGTGAAGCCCGCGGGGCGAGACCCGCTGGTGCATC 1202
453 .....ProGluProAlaProGly.....G 460
1201 ACCGATGACGCCCTCGCCCGGAGGTGCTGC..... 1170
460 InProAlaThrProAlaArgSerGlyAlaAlaAspProAlaAspProVal 476
1169 .....CGATCCCGGTTGTTGAAGACCCGACCTCGCCCGCGCG 1129
477 GlyHisProAlaAlaProAlaProGlyProGluProAlaProGlyThrArg 493
1128 CCGTGCAGGCGGTGAGACGAGCTTCACATCCCGTTCGAGCTGGCT 1079
493 uGlnProAlaThrProArg.....ArgSerGlyAlaAla 505
1078 CCGTTCACGCTCATCGCGTGGAGCGGAGCCACCGCGCTGCGCG 1029
505 sProAlaAspProValGlyHisProAlaAlaProAlaProGlyPro 521
1028 CATCAGCAGCACTGGCTCAACCCCGCGGCTGGCGGAGCGGAGATC 979
522 GluProAlaThrArgLeuGlnProAlaThrProAlaArgSerGlyAlaAl 538
978 GCATCGCGCGCAT.....CGCGGCGCGCTGCTCAGCGCACT..... 942
538 aAspProAlaAspProValGlyHisProAlaAlaProAlaProGlyP 555
941 .....CGCGAGCGCTCGCGCGGCTC.....GGCCAAACGCGCGA 906
555 roGluProAlaThrArgLeuGlnProAlaThrProAlaArgSerGlyAla 571
905 GGTGATGGCGGCTTCGGTACCACCTCCCGTGTGGTGCATCTGCGAC 856
572 AlaAsp.....ProAlaAspProValGlyHisProAlaAl 583
855 TGCCTGCTGTGGCGGTACACCGATCCGCGGATGGCCGCGAGCGCTCAGC 806
583 aProArg.....AlaProGlyProGluProAlaProGlyTha 594
805 GTTCTCAAGCACTCGCGCTCGCGG.....CCCGACGAG 771
594 rgLeuGlnProAlaThrProAlaArgSerGlyAlaAlaAspProAlaAsp 610
770 CGCGGGGGTGAAGGACGAGACCTGCGCGGGGCGTCCGCGGACACTCGG 721
611 ProValGly.....HisProAlaAlaProAlaGala..... 620
720 CCTGGAGAGCTGCTCTCGAAGCGGTGCATCGACCGCGGAGAACGAC 671
621 ..ProGly..ProGluProAlaThrArg..LeuGlnProAla..... 632
670 ACCCGGACCATGACCGCGCTGTGATGACGCGCGGAGCGGAGTTGCG 621
633 .....ThrProArg.....Arg 636
620 CTCGGTCTCCGACGACAGCTGCTACATGATCAGCGGCTCATCTTGG 571
637 SerGlyAlaAlaAspProAlaAspProValGlyHisProAlaAlaProAr 653
570 CGGAGCA.....CGACACGAGCGGCTCTCTCGGCTTCGCTGC 530
653 gAlaProGlyProGluProAlaThrArgLeuGlnProAlaThrProAla 670
529 GCGAGGTCTCGCGGCGG.....CCT 507
670 rgSerGlyAlaAlaAspProAlaAspProValGlyHisProAlaAlaPro 686
506 GCG.....GGCGATGCGGAGAGAGACCGCTCCCGCTT.....CG 469
687 ArgAlaProGlyProGluProAlaThrArgLeuGlnProAlaThrProAr 703

```

```

468 TCGAGAGCGCTGCGCTACACCCGCGGCTGCCTACAGCTTGTGAGG 419
703 gArgSerGlyAlaAlaAspProAlaAspProValGlyHisProAlaAlaP 720
418 TTGCTGCCACGAGGTGACATCGCGCGCT.....CGGCTGCGCGG 375
720 roArgAlaProGlyProGluProAlaThrArgLeuGlnProAlaThrPro 736
374 CGGAGCGCGGTGCTGT.....GCACATCG..... 349
737 ArgArgSerGlyAlaAlaAspProAlaAspProValGlyHisProAlaAl 753
348 ...AGGCAACCAACACGAGCGCGCCATCAGACGACCCCGACGCTT 303
753 aProAlaProGlyPro.....GluProAlaThrArgL 765
302 CCACCGGACCGTCCCTCGTGGCGGCTGCATCTTGGCGAGCGGCGCG 253
765 euGlnProAlaThrProAlaArgSerGlyAlaAlaAsp..... 777
252 ACTACTGCATCGGGAGAGCTGCGCCACGTGAGTCCGGCAGATGATC 203
778 ...ProAlaAspProValGlyHisProAlaAlaProAlaGala..... 790
202 GCGTACTGCGCAGCAGTTCCCGAGCGCGCAGCTGGCGCTGCCATAGA 153
791 .....ProGlyProGlu..... 794
152 CGAGTTGCGGTGCGGAGAGGCGCCAGACGCGCGCTCACCAGAC 103
795 .....ProArgThrArgLeuGlnProAla 802
102 TGCCTG.....TTCGCTGCGCTGATCCGCGCGCGGAGCGGAGCGG 59
803 ThrProAlaArgSerGlyAlaAlaAspPro.....AlaAspProValG 817
58 ACCGCGCAGCGGAGCAGCGCGGTGGCGGCTCCCGTCCGCTATCCGG 10
817 yHisProAlaAlaProAlaGala...ProGlyProGluProAlaThrArg 832

```

seq_name: p1r2:702345

seq_documentation_block:
hypothetical protein KIA00324 - human (fragment)

C:Species: Homo sapiens (man)

C>Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 05-Nov-1999

C:Accession: T02345

R:Rickey, D.O.; Bruce, D.; Mundt, M.; Doggett, N.; Munk, C.; Saunders, E.; Robinson, D.

re, J.; White, S.; Yeng, S.; Fatum, O.; Campbell, C.; Fawcett, J.; Deaven, L.

submitted to the EMBL Data Library, March 1998

A:Description: Sequencing of human chromosome 1p13.3.

A:Reference number: Z14664

A:Accession: T02345

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1791 <RIC>

A:Cross-references: EMBL:AC004493; NID:g2996648; PIDN:AA08453.1; PID:g2996650

C:Genetics:

A:Map position: 16

A:Introns: 1610/2; 1706/2

A>Note: KIA00324

alignment_scores:
Quality: 486.50 Length: 1108
Ratio: 0.977 Gaps: 55
Percent Similarity: 44.946 Percent Identity: 27.978

alignment_block:

us-09-673-254-1/rev x 702345 ..

Align seg 1/1 to: T02345 from: 1 to: 1791

```
2844 CCAGCAACCGCGCG.....GCATGTTCTGGTCGGCACACGACCCAG 2801
    ||| |||||::: |||||::: |||
464 ProArgThrProSerArgSerArgSerArgSerProGluLeuArg 480
    ||| |||||
2800 GAGGGTGTGAGGCCACATCCAGGTCATCAGCTCGAGCGCTTCTGCT 2751
    ||| |||||
480 GAspGlySerGlyThrProSerArg..... 488
2750 GGCAGGCTGCTGCGGAGCGGCTCGCGGCGGCGGCTGATCCTCAGCT 2701
    |||||::: |||||::: |||
489 ..HisSerLeu.....SerGlySerSerProGlyMetLysAspLeuPro 502
2700 CGTCCAGCGGTACACCCAGGCGGATGACCGCGAGCACATCAACGCG 2651
    |||||::: |||::: |||
503 .....ArgThrProSerArgGlyArgSerGlyLysAspSerSer... 515
2650 GACCGTCACCGCTACAGCGCGCGCTACGCGACGTCGAACAGCGC 2601
516 .....ProGluProLysAlaLeuProGluThrP 525
2600 CAACATCATGACCGCGCGAGCGCCGACGCGCT.....GGCGCG 2560
    ||| ||||| ||| |||||::: |||
525 roArg.....ProArgSerArgSerProSerSerProGluLeuAsn 539
2559 ACGTGCTGGCGGTACGCTATCACCCTGGGTAGTCCGACCCGATCGGA 2510
    |||::: ||||| |||::: |||
540 LysCys.....LeuThrProGluArgGlyLysSerGlySerGly 552
2509 CGGG.....GCAGCGTGCCT.....CGTC 2490
    :::: |||||::: |||
552 userSerValAspGluThrValAlaArgThrProLeuGlyGluArg 569
2489 CTACTTCGGTCAAC.....CCTTCCTGCGCTCG 2458
    |||||::: |||||::: |||
569 erArgSerGlySerSerGluLeuAspValLysProSerAlaSerPro 585
2457 CGGCGAAGGGCGCGACACCTCGTGGTGGCGTCCGCGCGCGCGCAG 2408
586 GlnGluArgSerLeuSerAspSerProAspSerLysAlaLysThrArg 602
2407 GAGTTGACACGAGGCGGCTACTACAGCGAGCGGCGGCTGCCGAGT 2358
    |||::: |||::: |||::: |||
602 gThr..ProLeuArgGluArgSerArgSerGlySerSerProGluVala 618
2357 CGGCGCGACGCGCG...ACGCGGCGCTCGCGCGAGACTGTGGAGCGCG 2311
    :::: |||::: |||||::: |||
618 sPerLysSerArgLeuSerProArgArgSerArgSerGlySerSerPro 634
2310 .....GCGGCGCGCGTGGGAGACGCGGACGACTGAGCGGCGCGC 2269
635 GluValLysAspLysProArgAlaAlaProArgAlaGlnSerGlySerAs 651
2268 .....CCGCCCGCGCGCATGTCGTTCTCATCCGAGAGATG 2232
651 pSerSerProGluProLysAlaProAlaProArgAlaLeuProArgArg 668
2231 CGCTTCATCCGAGCGCGACGCTGCTGCCGATCATTCGAAGGAA 2182
    |||::: |||||::: |||::: |||
668 erArgSerGlySerSerLysGlyArgGlyPro..SerProGluGlySe 684
2181 CGATTGTCAGCAGGTGCGCGCGCGCGCGCGCATGTCGACGCTCGG 2132
684 rSerSerThrGluSerSerProGluHisProProLysSerArgThrAla 701
2131 TTGCGC...CCGATGTCGCGCTCGCGCGGACTTCTACACCGGCGCTGTCG 2085
    ||| |||||::: |||::: |||
701 rghArgGlySerArgSerSerProGluProLysThrLys.....Ser 714
2084 GCTGGACCGCGACCGGTGTGACGACCGCGCGCGCGGAGATACAGACG 2035
    |||||::: |||||::: |||
715 ArgThrProProArgArgSerArgSerSerProGluLeuThrArg 731

2034 TTCAGTCCGAGCGGAGAGCTGGTGGTCCGCGGTCGCCGACCATGCA 1985
    |||||::: ||| |||||::: |||
731 gLysAla.....ArgLeuSerArgArgSerArgSerAlaSerSer 745
1984 CACCCCTACACCGCGCTACGCGCGCGGAGAACGACGACGAGCGCATCG 1935
    ||| |||::: |||||::: |||
745 erProGluThrArgSerArgThrProProArg..... 755
1934 CGGCAATCTGACCGGTACTTCGCCACCGACGACGCGGACGACTGACC 1885
    |||::: |||::: ||| |||
756 HisArgArgSerProSerValSerSerProGluProAlaGluLysSer 772
1884 AAGCGGTGACAGCGCGCGCGCGGCGGAGTCATCATGACTCCGATGAC 1835
772 gSerSerArgArgArgArgSerAlaSerPro.....ArgThrLysT 787
1834 CCTGGCGCTCGCGCGATGCGGCTTTCGCGACCGCGCGCGCG...CCG 1788
    |||||::: ||| ||||| ||| |||
787 hrThrSerArgGlyArgSerProSerProLysProArgGlyLeuGln 803
1787 CGTTCGCGCT...GGCGAAGGAGTCATGAGGCGCGGAGTGAGC 1741
    |||||::: |||::: |||||::: |||
804 ArgSerArgSerArgSerArgGlyLysThrArgThrArgArgArg 820
1740 GCGGTCCCGCGCTCGGCTGGGTGCGAGCTGTCGACCGGACGCTCGG 1691
820 GAspArgSerGlySerSerGlnSerThrSerArgArgGluArgSer 837
1690 GCGCGCGCGGACTTCTACCGCGCGACCTCGCGCTCGGCGCGGACA 1641
    ::: |||||::: ||| |||||::: |||
837 rSerArgSerArgValThrArgArg..ArgArgGlyLysSerGlyThr 853
1640 C.....CGGACTGAAGGCGCTCACC 1621
853 sSerArgSerProAlaArgGlnGluSerSerArgThrSerSerArgArg 870
1620 GACCCGCTGCGACATCGGTGACACACCGCGTCCGCGGACCC..AGGAC 1572
    |||::: |||::: |||||::: |||
870 rghArgLysArgSerArgThrProProThrSerArgLysSerArgSer 886
1571 TGGCGCTACCGCGCGGTACGCGCGACCTGGCGCTGCTTCGCGCGT 1522
    :::: |||||::: |||::: |||
887 ArgThrSerProAlaProThrLysArgSerArgSerArgAlaSerPro 903
1521 CAGGACTGCGACGCGAGCGTCCGCGCGCGCTGACACTCG.....GCGG 1478
903 atThrHisArgArgSerArgSerArgThrProLeuLysSerArgArgArg 920
1477 CTCGCTCGAGAACGACCGCGCGACACGCGCGGCGGCGGCGGAC 1428
    :: |||||::: |||||::: |||
920 erArgSerArgThrSerProValSerArgArgArgSerArgSerArgThr 936
1427 TGCTGACCGCGACGCGGCGGCTTCTCGGTGTCGACCTCGGAGGG 1378
    ||||| ||||| |||::: |||
937 SerValThrArgArgArgSerArgSerArgAlaSerProValSerArg 953
1377 TACCGCGCGCGCGGCGGCTCGCTGAGCGGCGAGG...CGCCCG 1332
953 gArgSerArgSerArgThrProProValThrArgArgArgSerArgSer 970
1331 GGTGGCGTGCAGCCCTTCGCGGTCGATGATGACCATGACAGCGGACG 1282
    |||||::: ||| |||
970 rghThrProThrArgArgArg..... 977
1281 CCGAGGTGACAGCGCTTCGCGGAGCGGCGCGGCTCGAGGTGAA 1232
    ||| ||||| ||| ||||| |||
978 SerArgSerArgThrProProValThrArgArgArgSerArgSerArg 994
1231 GCGCCCG.....CGGCGAGACCGCGCTGGGTGTCATCAGCATGACG 1191
    |||||::: |||::: |||||::: |||
994 rProProValThrArgArgArgSerArgSerArgThrSerProLysThr 1011
1190 CTTGCGCGCGAGAGTGTGCGCATCCCGGTCTGTGAAGACCGCGAC 1141
```

```

1011 rgarargSerArg..SerArgThrSerProValThrArg...ArgArgse 1026
1140 TCCGCCCGCCGCTGGCGGGGGTGAGCAGCGTCTCGACATCCCGCTT 1091
1026 rArgSerArgThrSerProValThrArgArgArgSer.....ArgS 1040
1090 CCGGAGCTGCGCTTCACGCTCATCGCCGTCAGCGGCGAGGCCACCG 1041
1040 eArgThrSerProValThrArgArgArgSerArgSerArgThrPro 1056
1040 GCGGCTGCGCGCATCCAGCAGCTGCTCAACCGCGCGCGCTGGCGG 991
1057 AlaAlaArgArgSerArgThrProLeuThrProAlaAla 1073
990 AGGGAGCGATGCGATCGCGGATCGCGCGCTGCTCACCAGATC 941
1073 gSer.....ArgSerArgSerProLeuAlaAlaArgArg 1085
940 GCGGAGCTGCGCGCTGCGCGCAACCGCGAGCTGATCGCGGCTT 891
1085 rArgSerArgSerArg...ThrProArgThrAlaArgGlyLysArgSerLeu 1100
890 CCGGTACACACTTCCGCTGTGTCATCTGAGACTGCTGCTGTCGGG 841
1101 ThrArgSerProAlaAlaArgArgSerAlaSerGlySerSerSe 1117
840 TCACGCGATCGCGGATGCGCGGAGCGCTCAGCGCTTCACAGGCTC 791
1117 rAspArgSer..Arg...SerAlaThrProProAlaThrArgAsnHisSe 1132
790 GCGCTGCGCGCGCGCGAGCGCGGGGTGACGCGCAGCAGCTTGGCG 741
1132 rGlySer.....ArgThrProProVal 1139
740 GCGGCTGCGCGAGCTGCGCGCTGAGAGCTGCTCTCGAAAGCG... 694
1139 alaAlaLeuAsnSerArg.....MetSerCysPheSerArgProSer 1153
693 .....TGCACTCAGCGCGCGCGGAGACGACGACG 668
1154 MetSerProThrProLeuAspArgCysArgSerProGlyMetLeuGlyPr 1170
667 CCGGACATGACCGCGCTGTACGAGCGCGCGCAGTTCGGGCTC 618
1170 o.....LeuGlySerSerArgThrProMetSerVal 1181
617 GGTCTCCGAGCAGCAGCTGTCATATGATCACCAGGCTCATCTTCCG 568
1181 euGlnGlnAlaGlyLysMetMetAspLysProGlyProAlaGlyLePro 1197
568 ..... 568
1198 AspHisGlnArgThrSerValProGluAsnHisAlaGlnSerArgTleAl 1214
567 .....GCCAGCAGCAGCGGCTCTCTCTCT 543
1214 alaAlaLeuThrAlaLeuSerLeuGlyThrAlaArgProProSerMet 1231
542 GGGCTTCTCTCTGCGGAGAGTCTGCGGGGCGCGCGCGAGATCCG 493
1231 etSerAlaAlaGly.....LeuAlaAlaArgMetSerGln 1242
492 ACAGAGGAGCGCGCTCTCCGCTGTGAGAGAGCGCTGCGCTACCA 443
1243 ValProAlaProValPro.....LeuMetSerLeuArg 1253
442 CCGGTGCGCTACACGTTGTGAGAGTTCGTCACGAGGAGACCATCG 393
1253 gThrAlaPro.....AlaAlaAsnLeuAlaSerArgTlleProAla 1267
392 CCGGCTCGGGCTGC.....CCCGCGAGCGC 367

```

```

1267 laSerAlaAlaAlaMetAsnLeuAlaSerAlaArgThrProAlaTlePro 1283
366 CGGCTGTGTGACATGAGGAGGAGCAACCGGAGCGCGCG... 325
1284 ThrAlaValAsnLeuAlaAspSerArgThrProAlaAlaAlaAlaMe 1300
324 .....ATCAGCAGCGCGCGCAGC..... 307
1300 taSnLeuAlaSerProArgThrAlaValAlaProSerAlaValAsnLeu 1317
306 .....CCTTCACCGCGGAGCTCTCT.....CTGCGCGCGCG 274
1317 laAspProArgThrProThrAlaProAlaValAsnLeuAlaGlyAlaArg 1333
273 TCACCTGCGGAGAGCGCGCGCAGTACGTCAGTTCGCGGAGCGCTG 224
1334 ThrProAlaAlaLeuAlaAlaLeuSerLeuThrGlySerGlyThrPro 1350
223 CTGAGTTCGCGCAGATGATCGCGCTACTGCGCAGCAGTTCGCGAGC 174
1350 oThr.....AlaAlaAsnTyrProSerSerSerArgThrPr 1362
173 CCGACTGCGCGCTGCGCTGACGAGAGTTCGCGT..... 142
1362 roGlnAlaProAlaSerAlaAsnLeuValGlyProArgSerAlaAlaAla 1378
142 ..... 142
1379 ThrAlaProValAsnTleAlaGlySerArgThrAlaAlaAlaAlaPr 1395
141 .....GTGCGCGAGAGGCGCGCAGAGCGCGCGC..... 112
1395 oAlaSerLeuThrSerAlaArgMetAlaProAlaLeuSerGlyAlaAsn 1412
111 .....TCACGAGACTGCGCG.....TGTGCTGC 88
1412 euThrSerProArgValProLeuSerAlaTyrGlnArgValSerGlyArg 1428
87 GGTGATTCGCGCG.....GCCGCGACCGCGCAGCGGAGCGCGC... 52
1429 ThrSerProProLeuLeuAspArgAlaArgSerArgThrProProSerAl 1445
51 .....ACCGCGCAGCGCGCGGTGCGCGCTCCGCTGC 21
1445 aProSerGlnSerArgMetThrSerGlnArgAla.....ProSerP 1459
20 CGTGTACCGCGTGCGA 4
1459 roSerSerArgMetGly 1464
seq_name: p1r2:T02345
seq_documentation_block:
hypothetical protein KIAA0324 - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #extl_change 05-Nov-1999
C:Accession: T02345
R:Ricke, D.O.; Bruce, D.; Mundt, M.; Doggett, N.; Munk, C.; Saunders, E.; Robinson, D
re, J.; White, S.; Ueng, S.; Tatum, O.; Campbell, C.; Fawcett, J.; Deaven, L.
submitted to the EMBL Data Library, March 1998
A:Description: Sequencing of human chromosome 16p13.3.
A:Reference number: Z14664
A:Accession: T02345
A:Status: preliminary, translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1791 <RIC>
A:Cross-references: EMBL:AC004493; NID:g2996648; PIDN:AA08453.1; PID:g2996650
C:Genetics:
A:Map position: 16
A:introns: 1610/2; 1706/2
A:Note: KIAA0324

```

alignment_scores:
 Quality: 479.00 Length: 1056
 Ratio: 1.028 Gaps: 56
 Percent Similarity: 44.129 Percent Identity: 29.261

alignment_block:
 US-09-673-254-1 x T02345 ..

Align seg 1/1 to: T02345 from: 1 to: 1791

```

133 TCCGCGACACCGCACTGCTGTACGGCAGCGCCATCTC.....GG 173
      ::::: ::::: ::::: :::::
426 AAGlyMetSerSerSeraInglSerSerProValLeuAspAlaYa 442
      ::::: ::::: ::::: :::::
174 GCGTCGGGAGAACCTCGTCGGCAGTACGCCGATCATCTGCGGACTCCAG 223
      ||||| ||| ||| ||||| :::::
442 LProArg..ThrProSerArgLuarSerSerSeraAlaSerSerProg 458
      ::::: ::::: ::::: :::::
224 CTGGGCGA...GCTGCTCCCGATGCAGTACGTACGGCCGCTCGCCGAGAG 270
      ::::: ::::: ::::: :::::
458 LMetLysAspLLeuProArgThrProSerArgArgSerArgSergLy 474
      ::::: ::::: ::::: :::::
271 TGAGCCGCGCCGACGAGGAGCGTCCGGT..... 300
      ::::: ::::: ::::: :::::
475 SerSerProGlyLeuArgAspLysSerGlyThrProSerArgHisSerLe 491
      ::::: ::::: ::::: :::::
301 ...GGAAGGCGTGGCGGCGCTGCTGATGCGCGCGT..... 333
      ||| ||| ::::: |||
491 uSerLysSerSerProGlyMetLysAspLLeuProArgThrProSerArg 508
      ::::: ::::: ::::: :::::
334 ..CGGTGTGTGTGCTCCCTCGATGTCCACGACACCGGCGCTCGCGGGGCA 381
      ::::: ::::: ::::: :::::
508 LArgSerLLeuLysAspSerSerProGluProLysAlaLeuProGluThr 524
      ::::: ::::: ::::: :::::
382 GCGGAGCGCGCGCGATGTCACCTCGCTGGCAGCAGCACTCCACACGCG 431
      ||| ||||| ::::: :::::
525 ProArgProArgSerArgSerProSerSerProGluLeuAsnAsnLysGly 541
      ::::: ::::: ::::: :::::
432 TAGGGACCG.....GCGGGTGTAGCGCAGCGCTCTCTCCAC 469
      | ::::: ::::: ::::: :::::
541 sLeuThrProGluArgGluArgSerLysSerLysSerValAspGluL 558
      ::::: ::::: ::::: :::::
470 GAACGCGGAGACGGCGCTCTCGTGGCATCCGCGCGAGCGGC..... 513
      ::::: ::::: ::::: :::::
558 yethrValAlaArgThrProLeuGlyLnaArgSerArgSergLysSer 574
      ::::: ::::: ::::: :::::
513 ..... 513
575 GluGluLeuAspValLysProSerAlaSerProGluGluArgSerLys 591
      ::::: ::::: ::::: :::::
514 .....CCGCCAGACCTCCGCGAGCAGAGCCAGAGCCAGAGAGAGCC 553
      ||| ::::: ::::: ::::: |||||
591 rAspSerSerProAspSerLysAlaLysThrArgThrProLeuArgLna 608
      ::::: ::::: ::::: :::::
554 GGTGTGTGTGTGCGCGCGAGATGAGCCCGGTGATCATGTAGACAGCT 603
      ||| ::::: ::::: :::::
608 rSerArgSergLysSerSerProGluValAspSerLysSerArgLeuSer 624
      ::::: ::::: ::::: :::::
604 GGTGCGCGAGACCGGAGCCGAACTCGGCGCTCG.....CGCGCTCG 644
      ||||| ||| ::::: |||
625 ProArgArgSerArgSergLysSerProGluValLysAspLysProArg 641
      ::::: ::::: ::::: :::::
645 TACAGACACCGGGTCATGTGCGGGGTGCTGCTCCGCGCGGCTAGTGAC 694
      ::::: ::::: ::::: :::::
641 gAlaAlaLProArgAlaGlnSerLysSerAspSer..... 652
      ::::: ::::: ::::: :::::
695 GCGTTCGAGAGACGGCTCTCCAGGGCCGAGGTGTCGCGACGCCCCCGG 744
      ||||| ||| ::::: |||||
745 CAGGTCGCGCGGTCAACCCCGCGCTCTGCGGCGCGCGAGCGAGT 794
      ||| ||||| ::::: :::::
663 AlaLeuPro...ArgArgSerArgSergLysSerSerLysLysArg 678

```

```

795 GCGTTGAGAACGCTGACGGCCTTCGCGG.....CCATGCG 829
      ||| ::::: |||
678 yPro.....SerProGluGlySerSerSerThrLysSerSerp 691
      ::::: ::::: ::::: :::::
830 CGGATCGGTGACCGGCACACGAGCAGCTCGCAGATGACCAACAGCGGGA 879
      ::::: ::::: ::::: :::::
691 roGluHisProProLysSerArgThrAlaArg..... 702
      ::::: ::::: ::::: :::::
880 AGTGTACGCGAAGCCCGCATCAGCTCGCGGCTTGGCCGACGCGCG 929
      ||||| ||| ::::: |||
703 .....GlySerArgSerSerProGluPro..... 710
      ::::: ::::: ::::: :::::
930 GAGCGCTGCGCGATGTCGTGAGACACCGCGCGCATCGCGCATCGC 979
      ::::: ::::: ::::: :::::
711 ...LysThrLysSerArgThrProProArgArgArgSerArgSers 726
      ::::: ::::: ::::: :::::
980 ATCCGTCCTCGCGCCAGCGCGCGGCGTTGAACGAGTGGGTGGATGC 1029
      ::::: ::::: ::::: :::::
726 erProGlu..LeuThrArgLysAlaArgLeuSerArgArgSerArgSera 742
      ::::: ::::: ::::: :::::
1030 GCGCAGCGCGCGGTGGGCTCGCGCTCCACGCGCATGAGCGTGAACGGA 1079
      ::::: ::::: ::::: :::::
742 aSerSerSerPro.....GluThrArgSerArgT 752
      ::::: ::::: ::::: :::::
1080 GCGAGCTCCGGAACGGGATGTCGAGACCGTGT...CCACCCCGCGC 1125
      ::::: ::::: ::::: :::::
752 hProProArgHisArgArgSerProSerValSerSerProGluProAla 768
      ::::: ::::: ::::: :::::
1126 .....AGCGGCGCGGGCGAGTGGGGGTCCT..... 1152
      ::::: ::::: ::::: :::::
769 GluLysSerArgSerSerArgArgArgSerAlaSerSerProArgTh 785
      ::::: ::::: ::::: :::::
1153 .....TCACGAACCGGGATGCGCCAGCACCTCGCGGCGAGAGCG 1192
      ||||| ||| ::::: |||||
785 rLysThrThrSerArgArgLysArgSerProSerProLysProArgGlyL 802
      ::::: ::::: ::::: :::::
1193 GTCATCGTGTGATACCCAGCGGGGTGCGCCGCGGGGGCGTTCACTGCA 1242
      ::::: ::::: ::::: :::::
802 euGlnArgSerArgSerArgSerArgGluLysThrArgThrArg 818
      ::::: ::::: ::::: :::::
1243 CAGACGCGGCGCGCTCCGGAAGCGCTGTCGACACTCGGGCTGCGCTCG 1292
      ||| ||| |||
819 Arg.....ArgAspArg..... 822
      ::::: ::::: ::::: :::::
1293 ATGCTCATCATGAGACACGACGAGCGGTGACGCGCACCGCGGGCGCTTC 1342
      ::::: ::::: ::::: :::::
823 ...SerGlySerSerGlnSerThrSerArgArgArgLnaArgSera 838
      ::::: ::::: ::::: :::::
1343 GCGCTCAGAGGACCGCGCGCGCGCGGG.....TACCCCTC 1383
      ||||| ||| ||||| |||
838 erArgSerArgValThrArgArgArgArg..GlyLysSerLysThrHis 854
      ::::: ::::: ::::: :::::
1384 CCGC.....AGTTCGACACCGAGAAACCGCGCCCGT 1415
      ||||| ||| ||||| |||
1416 GCGGTCGAGACAGTCGCGCGCGCGCGCGCTGGCGGTGCGGGGCTCG 1465
      ||||| ::::: ||| ::::: :::::
871 ArgGlyArgSerArgThrPro.....ProThrSerArgLysSerArg 885
      ::::: ::::: ::::: :::::
1466 TTCTCGACGAGC.....CGCCGATTCGAACGCGCGCGCGAGC 1503
      ||||| ||| ::::: :::::
885 gSerArgThrSerProAlaProThrLysArgSerArgSerArgAlaSer 902
      ::::: ::::: ::::: :::::
1504 CGTG.....GTCGCGAGTCGTCGAGCGGAGAACAGACGCGCAGTGGC 1547
      ||::: ||| ::::: |||
902 roAlaThrHisAlaArgSerArgSerArg.....ThrProLeuIle 915
      ::::: ::::: ::::: :::::
1548 GCGCTACCGCGC...CGGTGACCGCCAGCTCTGAGTGCAGCGAGCGGT 1594
      ::::: ||| ||| ::::: |||
916 SerArgArgArgSerArgSerArgThrSerProValSerArg..... 929

```



```

1595 GTGTACCATGTGTCAGACCGGGGTGTCAGCCCTTCAGTCCGGTGC 1644
    ||| ::||| ||| |||||
930 .....ArgArgSerArgSerArgThrSerValThrArgA 941
1645 GCGCCGAGCCAGCGCGAGGTCCCG.....GTTAGACTCCGGG 1685
    || ::||| ||||| |||||
941 rArgSerArgSerArgAlaSerProValSerArgArgSerArgSer 957
1686 CGG...CCCCGATGCGGTGCACACGCTCCAGCCAGCCGAGCCGCG 1732
    ||| ||| ||| ||| ||| |||
958 ArgThrProProValThrArgArgSerArgSerArgThrProThr 974
1733 GCGACGCGCCCTCACTCCGCGCCCTCATGATCCCTTCGCGCAGCCG 1782
    ||| ||| ||| ||| ||| |||
974 rArgArgSerArgSerArgThrPro.....ProV 985
783 GAACGCGCGCGCGCGGGGTGCGGGAAGACCCGATCCGCGAGCCGA 1832
    ||||| ||| ||| ||||| ||| |||
985 aThrArgArgSerArgSerArgThrProProValThrArgArg 1001
1833 GGACGTCATCCGATCATGATGACCT.....CGCGCGCGCGCGTC 1873
    ||| ||| ||| ||||| ||| |||
1002 SerArgSerArgThrSer.....ProIleThrArgArgArgSer 1016
1874 TCGACCGCGTGTGTCAGTCGTCGCGTCGTCGCGAGACATACAGGT 1923
    ||| ::||| ||||| ||| |||||
1016 rArgThrSerProValThrArgArgSerArgSerArgThrSer... 1031
1924 CCAGATGCGCGGATCCGCTGTCGTCCCGGCGCGCTACGCGCGGT 1973
    ||||| ||| ||| ||||| ||||| |||
1032 .....ProValThrArgArg...ArgSerArgSerArgThrSer 1044
1974 GGTAGGGGTGTCATGTCGTGCGGGCGACCGCGCGACACTTCGCG 2023
    ||||| ||| ||| ||| ||| |||
1045 ValThrArgArgSerArgSerArgThrPro.....ProAlaIle 1059
2024 T.....CGAGCTGAACGTCTGTATCCCGCGCGCGCGGTCCGT 2064
    ||||| ||| ||| ||||| |||
1059 gArgArgSerArgSerArgThrProIleuLeuProArgLys..... 1072
2065 GACACGCTGGCGGTCCAGCCGACAGCGCGGTGTAAAGTCGGCG... 2111
    ||| ||| ||||| |||
1073 .....ArgSerArgSerArgSerPro 1079
2112 .....AGGCGGACATCGGGGAGACCGAGTCGAACCTGCGGG 2152
    ||||| ||||| ||||| ||||| |||||
1080 LeuAlaIleArgArgArgSerArgSerArgThrProArgThrAla 1096
2153 GCGCGCGCGCGGACACTGTGTA.....CGAATCGTTCCTTTCG 2190
    ||||| ||| ||| ||| ||| |||
1096 yLysArgSerIleuThrArgSerProProAlaIleArgArgSerAla 1113
2191 ATGGATCGGCACACGAGCTTCGCTCCGGATGAGACGACATCCG 2240
    ::||| ||| ||| ||||| ||| |||
1113 ercLysSerSerArgSerArgSerArgThrProProAlaThrArg 1129
2241 GGATGAGACGACATCGGGGCGGGCGCGCGCTCAGCGCGCGG 2289
    ::||| ||| ||| ||| ||| |||
1130 AsnHisSerGly.SerArgThrProProValAlaLeuAsnSerSer 1146
2290 .....TGTGCGCGACG.....CGGCGG 2307
    ||| ||| ||| ||| ||| |||
1146 eSerSerProSerArgProSerMetSerProThrProLeuAspArg 1162
2308 CGCGCGCTCCAGAGCTTCGCGCGAGCGCGCTCGCGGCGCGCGG 2357
    ||| ||| ||| ||| ||| |||
1163 ArgSerProGlyMetLeuGlnProLeuGly.....SerSerAr 1175
2358 CTCACCGGGGACCGCGCGGTGTAGTACCCCGGTGTCAAC 2407
    ||||| ||| ||| ||||| ||| |||
1175 gThrProMetSerValLeuGlnAlaGlyGlySerMetMetAspGly 1192
2408 CTCGCGCG..... 2415

```

```

1192 rGlyProArgIleProAspHisGlnArgThrSerValProGluAsnHis 1208
    || |||
2416 .....GCGCGAGCGCCAGCCACAGAGGTGCG 2444
1209 AlaGlnSerArgIleAlaLeuAlaIleuThrAlaIleSerLeuAl 1225
2445 GCGCCCTTCGCGCGGAGCGAGGAAGGGTTGACCGGAGTAGACGA 2494
    ||||| ||| ||| ||| ||| |||
1225 aArgProProProSerMetSerAlaIleGlyLeuAlaIleArgMetSer 1242
2495 GCGACGCTGCGCCGTCG.....ATCGGGGTGCGACTACCGGGGTGA 2540
    ::||| ||||| ||| ||| ||||| |||
1242 lValProAlaProValProLeuMetSerLeuArgThrAlaPro..... 1256
2541 TACCTACCGCCACAGACATCCGCGCGCTGCGCGCTCCCGCGGT 2590
    ::||| ||| ||| ||| ||| ||| |||
1257 .....AlaAlaAsnLeuAlaSerArgIleProAlaAlaSerAlaAl 1271
2591 CATGATGTGGCCTGTTTGACGTGCGGTACGCTGCGCGCGCTGTAG 2640
    ||| ||||| ||| ||| |||
1271 aMetAsnLeuAlaSerAlaArgThrPro..... 1280
2641 GGTGACGTCGCGCTGTAGGTCGTCGCGGTGATCCGCGCTGCTGTA 2690
    ||| ||| ||| ||| ||| |||
1281 .....AlaIleProThrAlaValAsn 1287
2691 GCGTCGAGAGGTGAGATACACCCCGCGCGAGCGCGCTCCGCGAG 2740
    ::||| ||||| ||||| ||| |||
1288 LeuAlaAsp.....SerArgThrProAlaAlaAlaAlaLame 1300
2741 CAGCGCTGCGAGGAGCGTGCAGGTGATGACTGATGATGGTGGCT 2790
    ::||| ||||| ||| ||| ||| |||
1300 tAsnLeuAlaSerProArgThrAla.....ValAla. 1310
2791 CGAACCCTGCTGCTGTCGTGTCGCGGACGACAGACATGCG... 2831
    ||||| ||| ||| ||| ||| |||
1311 .....ProSerAlaValAsnLeuAlaAspProArgThrProThrAlaPro 1325
2832 .....CGGCGTTCCTGCGCATGACATC 2854
1326 AlaValAsnLeuAlaGlyAlaArgThrProAlaAlaLeuAlaLeuSe 1342
2855 GATGCGCGG 2864
1342 rLeuThrGly 1345
seq_name: p1r1:B45344
seq documentation block:
probable nuclear antigen - suid herpesvirus 1 (strain Kaplan)
C:Species: suid herpesvirus 1
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 16-Jul-1999
C:Accession: B45344
R:Vileck, C.; Kozmik, Z.; Paces, V.; Schirm, S.; Schwyzer, M.
Virology 179, 365-377, 1990
A:Title: Pseudorabies virus immediate-early gene overlaps with an oppositely oriented
A:Reference number: A45344; MUID:91021039
A:Accession: B45344
A:Status: translation not shown
A:Molecule type: DNA
A:Residues: 1-1733 <VIC>
A:Cross-references: GB:M34651; MID:g334070; PIDN:AAA47471.1; PID:g334072
C:Superfamily: pseudorabies virus 1 nuclear antigen
alignment_scores:
    quality: 469.00    length: 1226
    ratio: 0.925      gaps: 72
    percent similarity: 41.354    percent identity: 29.364
alignment_block:
us-09-673-254-1 x B45344 ..

```

Align seg 1/1 to: B45344 from: 1 to: 1733

```
2 GATCCGACCGGGTA...CAGCGACGCGGACCGCCACCGCGCGTCCG 48
   ||| ||||| |||||
631 AspGlyAlaArgValGlnHisGly.....ArgAlaGlnLeuAla 643
49 GGTGGGCGTCCCGTCCGCGTCCGCGCGGATCAGCCAGCCAGACG 98
   ||| ||||| ||| ||||| ||||| |||
643 AlaProGlyProAlaHisGlyAlaLeuGlyLysArgValGlnAlaAsp. 659
99 GGCAGTTCCGTGACGCGCGCGTCCGCGCGCGTCCGCGACCAAGCA 148
   ||| |||
660 .....AlaAspValAsp 663
149 CTCGTGTAAGCAGCGCGAGTCCGCGCGTCCGCGG..... 181
   ||| ||||| ||||| ||||| |||||
664 ValValAlaProHisGlyArgAlaValAlaArgGlyProValLeuAspGly 680
182 .....GAACCTGTCCGCGAGTACGCGCGATCATCTGCGCGACTCC 221
   ||||| ||||| ||| ||||| ||||| |||
680 LglnHisAspGlyProAlaArgProAlaArgAlaGlnProAlaGln..... 695
222 AGCTGGGCGAGCTGCTCCCGATGCACTGCTGCGCGCGCGCGAAGT 271
   ||| ||||| ||||| ||||| ||||| |||
696 .....ValLeuHisGlyAlaGlyGlnAla 703
272 GACGCGCGCGCAGCA..... 286
   ||| ||| |||
704 GlnValProArgArgGlnGlnGlnHisProLeuGlyValGlnAlaAlaAs 720
287 .....GGAGCGTCCGCGTGAAGCGCTGCGCGCGCGCTGTA 323
   ||| ||| ||| ||||| |||||
720 PValGlyAlaProGlyProValProGlyProGlyValArgValAlaArg 737
324 TGGCGCGCGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 373
   ||||| ||||| ||||| ||||| ||||| |||
737 LglnLysAlaValGlyGlnGly.....GlnLysArgArgGlnAla 750
374 GCGGCGCGAG.....CCGACGCGCGCGATGCTGCTGCGCGCGCA 417
   ||||| ||||| ||||| ||||| ||||| |||
751 AlaAlaAlaArgValProGlyArgAlaArgGlyAlaLeuGlyLysLeuG 767
418 ACCTCCACAAGTGAAGCGACCGCGCGTGAAGCGCGCGCTCC 467
   ||||| ||||| ||||| ||||| ||||| |||
767 LysAlaGlnLeuLeuValGly...GlnArgValAlaGlnHisHisAla 783
468 AGCAACCGGACGCGCTCC.....GTGCGCATC 499
   ||| ||| ||| |||
783 LysValLeuGlyValGlyTyrLeuProHisProGlyLysAlaAlaAlaG 799
500 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 549
   ||| ||||| ||||| ||||| ||||| |||||
800 ArgGlyAlaAlaAlaArgGlyAspValArgGln..... 810
550 AGCGCGTGTGTCGTGCGCGCGAGATGAGCGCGCGTGAATCATGAG 599
   ||||| ||||| ||||| ||||| ||||| |||
811 ....GlyLysArgValGlnGlyLysArgGlyArgAlaProGlnPheGly 826
600 AGCTGCTC.....GTGCGACAC 616
   ||||| ||||| ||||| ||||| ||||| |||
826 SprLeuLeuValHisGlnGlyAlaGlnHisLeuGlyArgAlaValGly 842
617 CGAGCGCACTGCGCGTCCGCGCGCTGACAGACAGCGCGGTAT..... 661
   ||||| ||||| ||||| ||||| ||||| |||
843 GlnGlyArgLeuGlyLysProArgArgValGlyLeuAlaGlyLysAsp 859
662 .....GTGCGCGTGTGCTGTCGCGCGCGTGAAGTACAGC 695
   ||||| ||||| ||||| ||||| ||||| |||
859 AlaLysAlaAlaValAlaGlyArgGlyValLeuGlnHisGlyProGlnArg 876
696 GCTTCGAGGACGAGCTTC.....CAG 718
   ||| ||| ||||| ||||| ||||| ||||| |||||
```

```
876 LysProGlyProValValLeuGlyGlyGlyGlyGlyGlnArg 892
719 GCGCGAGGTTCGCGCAC.....GCGCC 741
   ||| ||||| ||||| ||||| ||||| |||||
893 GlySerGlyLysArgSerGlyProGlySerGlyAlaAlaLeuAlaPro 909
742 CGCG.....AGGTCGTGCGCGTCAACCGCGCG 770
   ||||| ||||| ||||| ||||| ||||| |||
909 OglYProProValLeuPheValAlaValAlaValAlaValProAlaG 926
771 CTCGTGCGCGCGCGAGCGCGAGTCCCTTGAACCTGACGCGCTCGG 820
   ||||| ||||| ||||| ||||| ||||| |||
926 LysLysArgAlaGlyGlnProLeuValLeuValAlaValProGlyAla 942
821 GCGCATCGCGCGGTGCTGACGCGCGACGCGACACTGCGAGTACCA 870
   ||| |||
943 GlyProGlyArg..... 946
871 ACAGCGGGAAGTGTACGCGAACCGCG.....GATCAGCTCGCGC 911
   ||||| ||||| ||||| ||||| |||||
947 .....AlaAlaLeuLeuAlaProLeuGly 956
912 GCTTCGCGCGCGCGCGCGAGCGTGGCGAGTTCGCTGACAGCGCGC 961
   ||||| ||||| ||||| ||||| ||||| |||
956 ArgProValArgAlaGlyGly.....GlyAlaGlyAla 968
962 GCGCATCGCGCGATGCTGCTGCGCGCGCGCGCGCGCGGTGA 1011
   ||| ||||| ||||| ||||| ||||| |||
969 GlyGlyAlaGlyGlnAla.....GlyLeuGlyAlaGlyAla.GlyLeu 983
1012 ACAGCGGTGCTGATGCGCGCA...GGCGCGTGGCGT..... 1050
   ||||| ||||| ||||| ||||| ||||| |||
983 LysAlaGlyAla.GlyLeuGlyAlaGlyLysAlaGlyProGlyAlaG 999
1051 .....CGCGTCCAGCGCGATGAGCTGACGCGAGC 1081
   ||||| ||||| ||||| ||||| ||||| |||
999 LysLysAlaGlyGlyAlaArgArgArgArgArgArgArgArgArg 1016
1082 CAGCT.....CGGAACGGGAGTGTGACAGCT.....CGT 1113
   ||| ||| ||||| ||||| ||||| ||||| |||
1016 LysAlaGlyLeuLeuGlyProGlyLysArgGlyAlaLysArgGlyLeu 1032
1114 CCACCCCGCGCGAGCGCGCGCGAGTCCGCGTTCACAGACCG 1163
   ||| ||| ||||| ||||| ||||| ||||| |||
1033 .GlyProGlyProArgGlyGly.....LeuGlyGlnProG 1044
1164 GATCGCGCGACACTCGCGCGCGAGCGCTCATGCTGATACCGAGC 1213
   ||| ||||| ||||| ||||| ||||| ||||| |||
1044 LysArgLysHisValGlyArgGlyGlnGlyLysArgGlyAlaGlyPro 1060
1214 GGTGCGCGCGCGCGCG...GTTCACCTGACGACGCGCGCGCTCC 1260
   ||||| ||||| ||||| ||||| ||||| |||
1061 GlyLeuAlaGlyAlaGlyProValHis..... 1069
1261 GGAAGCGTGTGACACTGCGCGTGCATGCTCATGCGACAC 1310
   ||||| ||||| ||||| ||||| ||||| |||
1070 .....AlaValAlaHisGlnArgGlnHisGlyAlaG 1080
1311 GCGACGGGTGACGCGCACCGCGCGCGCTCGCGCTACAGACCGC 1360
   ||||| ||||| ||||| ||||| ||||| |||
1080 LysPheGlnLysP.ArgValArgGlyLeuProProLeuGlyArgAlaG 1096
1361 CCGCGCGCGCGGCTGACCGCTCCGCACTTGACACCGACGACCGCG 1409
   ||||| ||||| ||||| ||||| ||||| |||
1096 YPro.....GlyAspArgValAlaGlnArgGlnArgGlyArg 1110
1410 .....CCCCGTGCGGTGCA...CAGGTCCGCGC 1436
   ||||| ||||| ||||| ||||| ||||| |||
1110 LysLeuLeuLysAlaGlyGlyProGlnGlyLysArgGlyAlaGlyArg 1126
1437 GCGCGCGCTGCGCGTGTGCGCGCGCTGCTTCAGAGGACGCGCGAGT 1486
   ||||| ||||| ||||| ||||| ||||| |||
1127 Gly.GlnProGlyArgAlaGlyGlnGlnAlaLeuGlnLysAspAlaAlaG 1143
```

[illegible][illegible]

195 roAlaProGlySerThrAlaProAlaHisGlyValThrSerAlaPro 211
2586 CGCGGAGGCGCCAGAGCGCTGGCGGAGCTGGCGGTACGTATGAC 2537
212 AspThrArgPro.....Al 216
2536 CGCGGTAGCTCCGACCGCATCGAGCGGACGCTG.....C 2496
216 aProGlySerThrAlaProAlaHisGlyValThrSerAlaProAspT 233
2495 CTGCTCTACTCTCCGTTCAACCCCTCTCTGGCTCCGCGGAGAGCGG 2446
233 hArgProAlaProGlySerThr.....AlaProProAlaHisGly 246
2445 CGGACACCTGCTGTGGCTGGCGCGCGCGGAGAGGTGACCGACG 2396
247 ValThrSerAlaProAspThrAlaProAlaProGlySerThrAlaPro 263
2395 GCGC.....GCTACTACAGCGACCGGCGCTGCCCGGTAGCGGCC 2352
263 oAlaHisGlyValThrSerAla..... 270
2351 GACCGCGGACCGCGCTCCGCGGACGCTGGAGCGCGCGCGCGCG 2302
271 ..ProAspThrArgProAlaProGlySerThrAlaProAlaHisGly 286
2301 CGGTGGGACACGCGGCACT.....GACGGCGGCGCGCGCGCG 2261
287 ValThrSerAlaProAspThrArgProAlaProGlySerThrAlaPro 303
2260 CGCGGATGCTGCTATCCGCGAGATGCTGCTATCCGCGAGCGAG 2211
303 oAlaHis GlyValThrSerAla.....ProAspThrArgProAla 316
2210 AGCTCTGTGCTCCGATCCATCGAAAGAACGATGCTGACCGAGTTCG 2161
317ProGlySerTh 320
2160 CGCGGCGCGCGCGATGTTGACCTGGTGGCGCGGATGCGCGCTC 2111
320 ValaProProAlaHisGlyValThrSerAla...ProAspThrArgProAla 336
2110 GCGGCACTTCTACACCGCGCTGTTGCGTGGACCGCACCGTGTACG 2061
336 LaProGlySerTh.....AlaProProAlaHisGlyVal 347
2060 ACCGCGGCGCGCGGATACAGA.....GTTTCACTCCGAC 2023
348 ThrSerAlaPro...AspThrArgProAlaProGlySerThrAlaPro 363
2022 GGGAGCTGTGCGCGCGCTGGCGCGCGCATGACGACCGCTACCA 1973
363 oAlaHisGlyValThrSerAlaProAspThrArgProAlaProGlySer 380
1972 CGCGGCTAGCGCGCGGAGACGACGACGACGCA...TGCGGCGCATCT 1926
380 hAlaProProAlaHisGlyValThrSerAlaProAspThrArgProAla 396
1925 GACCGCTACTCTCCGACCGGACGAGCGGCGGACGACGACGACGAG 1876
397 ProGlySerThrAlaProProAlaHisGlyVal.....ThrSerAlaPro 411
1875 GAGAGCGCGCGCGGAGGTGATCATGACTCGATGAGCTGCTGGCGCT 1826
411 oAspThrArgProAlaProGlySer.....ThrAlaProProAla 424
1825 CGCGCGGATGCGGCTTTCGCG...ACCGCGCGGCGCGCGCTGCGCG 1779
424 LaHisGlyValThrSerAlaProAspThrArgProAlaProGlySer 439
1778 TGTGGCGCAAGGAGTCAATGAGGCGCGGAGGTACGCGGCTGCCGCG 1729
440ThrAlaProProAla 444

1728 TCGGTGCGCTGGGTGACCTGTGACCGACGACGACGCGCGCGGGA 1679
444 aHisGlyValThrSerAla.....ProAspThrArgProAlaProGly 458
1678 CTCTACCGCGGACCTCGGCTGGCTC.....CGG 1647
459 ..SerThrAlaProProAlaHisGlyValThrSerAlaProAspThr 474
1646 CGGACCGGAGTGAAGGCGCTCAGCAGCGGCTGTGACATCGGTGAT 1597
475 ProAlaProGlySerThrAlaProProAlaHisGlyValThrSerAla 491
1596 ACACCGGTGCGCGGACCGACGAGCTGGGCGCTGACCGCGGCTGCG 1547
491 oAspThrArgProAlaProGlySerThrAlaProProAlaHisGlyVal 508
1546 GCATGGGCGCTGTCTGCGCGCTGACGACTGACGACGCGAGCG... 1503
508 hSerAlaProAspThrArgProAlaProGlySerThrAlaProProAla 524
1502TCCGCGCGCGCTTGACTCGGCGCTCCGTCGACAGACG 1462
525 HisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAl 541
1461 CGCGCG.....ACAGCGCGGCGGCG 1439
541 aProProAlaHisGlyValThrSerAlaProAspThrArgProAlaPro 558
1438 GC.....GGCGGACCTGCTCGACCGCGAGG 1410
558 lSerThrAlaProProAlaHisGlyValThrSerAlaProAspThrArg 574
1409 CGGCTCTCGGTGCTGACTGCGGAGCGGTACCGCGCGCGCGCGG 1360
575 ProAlaPro.....GlySerThrAlaProProAlaHisGly 586
1359 GGTGCTGCTG.....GACGCGGAGCGCGCGCGGCTGCGCGAC 1317
586 yValThrSerAlaProAspThrArgProAlaProGlySerThrAlaPro 603
1316GTTGCGGTGCTCCATGATGAC 1296
1295 roAlaHisGlyValThrSerAlaProAspThrArgProAlaProGlySer 619
620 ThrAlaProProAlaHisGlyValThrSerAlaProAspThrArgProAl 636
1245 TCGTGAAGTGAAGCGCGCGGCGGACCGCGCTGGGTCAATGACGAT 1196
636 aProGlySer..ThrAlaProProAlaHisGlyValThrSerAlaProAs 652
1195 GACG...CGCTGCGCGCGGAGGTGCTGGCG..... 1168
652 rThrArgProAlaProGlySerThrAlaProProAlaHisGlyValThrs 669
1167 ..ATCCCGGCTCTGTAAGACCGCGGACGCGCGCGCGCTGCGG 1120
669 eAlaProAspThrArgProAlaProGlySerThrAlaProProAlaHis 685
1119 GGTGAGGACGCTC.....TCGACATGCC 1094
686 GlValThrSerAlaProAspThrArgProAlaProGlySerThrAlaPro 702
1093 G.....TTCGAGCTGCGCTGCTCAAGCTATGCGCGTGA 1056
702 oProAlaHisGlyValThrSerAlaProAspThrArgProAlaPro.Gly 718
1055 CGCGAGCGCGCGCGCGCTGCGCGGATCGACGACGACGCTGCTCAAC 1006
719 SerThrAlaProProAlaHisGlyValThrSerAlaProAspThrArgPr 735

OM of: US-09-673-254-1 to: SwissProt_40:* out-format : pfs

Date: Jun 11, 2002 8:21 PM

About: Results were produced by the Gencore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=framed_n2p.model -DEV=x1h
-O=/cgn21/USPTO.spool/US09673254/runtat_11062002.114213.1053/app_query.fasta.1.2973
-DB=SwissProt_40 -OPMT=fastan -SUPER=n2p.rsp -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOPCL=0.000 -LOPEXT=0.000
-OGAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-GOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blomsu62
-NS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct
-MAX=100 -THR MIN=0 -ALIGN=7 -MODE=LOCAL -OUTEMT=pfs
-HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09673254_@CGN1_1.80 -NCPD=6 -ICPD=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPPX -WAIT -THREADS=1

Search information block:

Query: US-09-673-254-1
Query length: 2870
Database: SwissProt_40:*
Database sequences: 105224
Database length: 38719550
Search time (sec): 57.220000

score_list:

Sequence	Strd Orig	ZScore	EScore	len	Documentation
SwissProt_40:VNUA_PRVKA +	468.50	285.30	1.9e-09	1733	P33485 pseudorabies virus (chick)
SwissProt_40:CA11_CHICK -	459.50	281.14	4.0e-09	1453	P02457 gallus gallus (chick)
SwissProt_40:CA11_CANPA -	459.50	281.14	4.0e-09	1460	P09xj7 canis familiaris (dog)
SwissProt_40:CA11_HUMAN +	457.00	279.62	4.8e-09	1464	P02452 homo sapiens (human)
SwissProt_40:MOC1_HUMAN -	455.00	275.73	5.5e-09	1255	P15941 h. mucin 1 precursor
SwissProt_40:CA11_MOUSE -	450.00	275.55	8.1e-09	1453	P11087 mus musculus (mouse)
SwissProt_40:CA13_MOUSE -	450.00	275.55	8.1e-09	1464	P08121 mus musculus (mouse)
SwissProt_40:MOC2_HUMAN -	449.50	267.09	6.7e-09	5179	P002817 homo sapiens (human)
SwissProt_40:YHL1_EBV -	442.00	275.91	1.7e-08	660	P03181 Epstein-Barr virus (str)
SwissProt_40:CA13_BOVIN -	442.00	272.94	1.6e-08	1049	P04258 bos taurus (bovine)
SwissProt_40:CA11_RANCA -	439.00	265.53	1.9e-08	1355	P04250 rana catesbeiana (bd)
SwissProt_40:CA25_HUMAN -	439.00	268.89	1.8e-08	1496	P05997 homo sapiens (human)
SwissProt_40:CA13_CHICK -	437.50	269.10	2.1e-08	1262	P12105 gallus gallus (chick)
SwissProt_40:VNUA_PRVKA -	437.50	267.07	2.0e-08	1733	P33485 pseudorabies virus (chick)
SwissProt_40:CF30_MYCTU -	434.50	277.49	3.5e-08	260	P03774 Mycobacterium tuberculosis
SwissProt_40:CA13_HUMAN +	432.50	265.20	3.0e-08	1466	P02461 homo sapiens (human)
SwissProt_40:CA11_HUMAN +	418.50	255.98	8.7e-08	1464	P02452 homo sapiens (human)
SwissProt_40:CA25_HUMAN -	417.00	251.61	8.6e-08	2944	P002388 homo sapiens (human)
SwissProt_40:CA15_HUMAN +	414.50	253.16	1.1e-07	1838	P20908 homo sapiens (human)
SwissProt_40:CA12_MOUSE -	413.50	254.06	1.3e-07	1459	P28481 homo sapiens (mouse)
SwissProt_40:CA13_HUMAN +	413.50	254.03	1.3e-07	1466	P02461 homo sapiens (human)
SwissProt_40:CA21_ONCMY -	412.00	253.65	1.4e-07	1356	P093484 oncomyoblastoma virus
SwissProt_40:CPXK_SACER -	411.50	261.12	1.9e-07	405	P33271 saccharopolyspora ery
SwissProt_40:CA17_HUMAN -	410.50	247.78	1.4e-07	2944	P002388 homo sapiens (human)
SwissProt_40:CA12_HUMAN -	409.50	251.89	1.7e-07	1418	P02458 homo sapiens (human)
SwissProt_40:CA11_CANPA -	409.50	251.70	1.7e-07	1460	P09xj7 canis familiaris (dog)
SwissProt_40:AMRH_YEAST -	408.50	251.24	1.9e-07	1367	P08640 saccharomyces cerevisiae
SwissProt_40:CPXH_STEGR -	406.50	255.07	2.7e-07	412	P26911 streptomyces griseus
SwissProt_40:CA21_MOUSE -	405.00	249.04	2.3e-07	1366	P08123 mus musculus (mouse)
SwissProt_40:CA13_MOUSE +	405.00	249.04	2.4e-07	1464	P08121 mus musculus (mouse)
SwissProt_40:CA13_BOVIN +	403.00	226.00	3.0e-07	1049	P04258 bos taurus (bovine)
SwissProt_40:CA15_HUMAN -	403.00	245.40	2.7e-07	1838	P20908 homo sapiens (human)
SwissProt_40:YHL1_EBV +	402.50	252.99	3.4e-07	660	P03181 Epstein-Barr virus (str)
SwissProt_40:CA12_MOUSE -	402.50	247.98	2.9e-07	1372	P01149 mus musculus (mouse)
SwissProt_40:CPXH_RHISN -	401.50	255.31	3.9e-07	400	P55544 rhizobium sp. (strain)
SwissProt_40:CA21_RAT +	399.00	245.93	3.8e-07	1372	P02466 rattus norvegicus (rat)
SwissProt_40:CA11_MOUSE -	398.00	244.97	4.1e-07	1453	P11087 mus musculus (mouse)
SwissProt_40:CA21_CANPA -	397.00	244.78	4.5e-07	1366	P04692 canis familiaris (dog)
SwissProt_40:CA1B_HUMAN -	397.00	242.98	4.3e-07	1806	P12107 homo sapiens (human)
SwissProt_40:CA11_CHICK +	396.50	244.09	4.6e-07	1453	P02457 gallus gallus (chick)

SwissProt_40:CA21_HUMAN + 396.00 244.19 4.8e-07 1366 P08123 homo sapiens (hum
SwissProt_40:CA1B_HUMAN + 395.00 241.81 4.9e-07 1806 P12107 homo sapiens (hum
SwissProt_40:CA11_BOVIN + 394.50 246.92 6.0e-07 779 P02453 bos taurus (bovine
SwissProt_40:CA21_BOVIN + 394.50 243.32 5.4e-07 1364 P02465 bos taurus (bovin
SwissProt_40:CA25_HUMAN + 394.50 242.72 5.3e-07 1496 P05997 homo sapiens (hum

seq_name: SwissProt_40:VNUA_PRVKA

seq_documentation_block:

ID	VNUA_PRVKA	STANDARD:	PRT:	1733 AA.
AC	P33485:			
DT	01-FEB-1994 (Rel. 28, Created)			
DT	01-FEB-1994 (Rel. 28, Last sequence update)			
DT	01-FEB-1994 (Rel. 28, Last annotation update)			
DE	Probable nuclear antigen.			
OS	Pseudorabies virus (strain Kaplan) (PRV).			
OC	Viruses; dsDNA viruses, no RNA stage; Herpesviridae;			
CC	Alphaherpesvirinae; Varicelloviruses.			
OX	NCBI_TaxID=33703;			
RN	[1]			
RP	SEQUENCE FROM N. A.			
RX	MEDLINE=91021039; PubMed=2171211;			
RA	Vitek C., Kozmik Z., Paces V., Schlim S., Schwyzer M.;			
RT	"Pseudorabies virus immediate-early gene overlaps with an oppositely			
RT	oriented open reading frame: characterization of their promoter and			
RL	enhancer regions.";			
RL	Virology 179:365-377(1990).			

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; M34651; AAA47471.1; -
PIR; B45344; B45344.
FT DOMAIN 112 117 POLY-THR.
FT DOMAIN 179 1733 GLY-RICH.
FT DOMAIN 192 196 POLY-SER.
FT DOMAIN 271 298 POLY-PRO.
FT DOMAIN 304 308 POLY-ARG.
FT DOMAIN 883 889 POLY-GLY.
FT DOMAIN 1398 1405 POLY-GLY.
SQ SEQUENCE 1733 AA; 172166 MW; 0C8CDBDE475B5E2 CRC64;

alignment_scores:

Quality:	469.00	Length:	1226
Ratio:	0.925 <td>Gaps:</td> <td>72</td>	Gaps:	72
Percent Similarity:	41.354	Percent Identity:	29.364

alignment_block:

US-09-673-254-1 x VNUA_PRVKA ..

Align seg 1/1 to: VNUA_PRVKA from: 1 to: 1733

2	GATCCGACCCGGGTA...CACGGACAGGGACCCGCGGTCGCGC	48
111		111
631	ASPGYALAAArgValGlnHisGly.....ArgAlaGlnLeuAla	643
111		111
49	GGTGGGCGGTCCCGGCGCGGTCCGGCGGATCAGGCGCAGAG	98
111		111
643	AAIAPRGGYPRoAlaHisGlyAlaLeuGlyArgValGlnAlaAsp	659
111		111
99	GGCAGTTCGATGAGCGCGCGCTGCGGCCCTTCGCGCACCCGCA	148
111		111
660AlaAspValAsp	663
149	CTGCTGTCGTCAGCGCAGCGCAGTCGCGCGG.....	181
111		111
664	ValValValPRoHisGlyArgAlaValArgGlyPRoValLeuAspGly	680

[illegible]

1203 LeuLeuGlyAlaIleValAlaAlaAlaHisGlyHisGlyAlaHisArgVala 1219
1674 AGAAGTCCCGGGCGCCCGCATGCCGTC.....GTCACACAGCTC...G 1714
1219 IArg.....GInGlyProGluArgValLeuGlyGlyHisGlyValProA 1234
1715 ACCGACCGACCGGACCGGCGCCGTCACCTCCGCGCCCTCCATGAC 1764
1234 sPValArgIArgIArgArgGlyHisAla..... 1242
1755 TCCCTGGCGCCAGACCGGACCGCGCCCGCGGCGTCCGGAAGACCG 1814
1243AlaGlyAspGlu.....GlyAlaValAlaValGlyArg 1253
815 C...CATCCGCGCCGAGCCCGGAGAGCTCATCGAGTCACTGATGACCTCG 1861
1253 yValAspProAlaLeuAlaGlyLeuValGlyAlaLeuValGlyGlyLeuH 1270
1865 CGCGCCGCGCTCCGACCGCGCTGGTCAGTGGCGGCGCTCGTGGTGGC 1911
1270 IsProArgValLeuGlnPro...GlyHis.....GlyLeuAlaGlyGly 1283
1911 1911
1284 ProAlaValAspGluAlaHisGlyLeuValLeuLeuProArgValArg 1300
1912GAAGTACAGGTCAGATGGCCGCGCATGCC... 1941
1300 HisLeuArgAspGluAspGlyHisGlyPro...GlyArgGlyAlaVala 1316
1942GTCTGTGTCTCCGCGCGCGCTACGCGCGC 1971
1316 IagIyArgGlyLeuAlaAspValAlaLeuValProGluProLeuAlaGly 1332
1972 GTGTAGAGGGGT.....GTCTGTGTGTGGCGGC... 2001
1333 ValProGlyAlaAlaValAlaAspAlaAlaValGlyArgArgValGlyAl 1349
2002GACCGCGCGACAGCTTCCCGTCGAGCT... 2031
1349 aGlyProGlyLeuProGluArgGlyGlyGlnArgProValGlyArgArg 1366
2032GAAGTCTG...GTATCCCGCGCGCGCGCGCT 2064
1366 IyProValGlyHisGlyArgGlyValValGlyGlyAlaAlaLeuPro 1382
2065 GACACAGGTGGCGGTCACGACCGAGCGCGGTGAAGTCCGCGAG 2114
1383 AlaArgGlyProGlyGlyLeuArgGlyArgGly...ArgGlyGlyArgGly 1398
2115 CGCGCAGAT.....CGGCGAAGCGAGTCCAGACCATGCGG 2151
1398 yGlyGlyGlyGlyGlyGlyGlyGlyArgGlyProArgArgArg.....GlyG 1413
2152 GCGCGCGCGCGACCTGTGTCAGCATGCTTTCATGATGATGCGGA 2201
1413 IyArgArgArgArgArgArgArgArgProGlyAlaGlyGlyIleArgGlyGly 1429
2202 CACGAGCGCTG.....CGCTCGCGGATGAG 2227
1430 ProAspSerPheValIlePhePheSerLeuGlyGlyGlyArgGlyArgGly 1446
2228 ACGGACATCTCCCGATGAGACGACGACATCGCGCGCGC...GCGCGCGCGC 2276
1446 yArgGlyGlyArgGlyGlyArgGlyArgGlyGlyGlyAlaProArg 1463
2277 CGTCACTGCGCGGTGTCGCGAGCGCGCGCGCGCTCCAGAGCTT 2326
1463 IyGlyGlyGlyGlyProGlyGlyGlyGlyArgAlaGlyArgGlyGlyVal 1479
2327 CGCGCGAGCGCGCGCTCGCGCGCTCAGCGGAGACCGCGC 2376

1480 ArgValAlaAlaAlaAlaAlaGlyAlaAlaGlu..... 1490
2377 GGTCCGTGATAGCGCGCCGTCGATCACTCTCCGCGCGG..... 2416
1491AlaAlaAlaAlaAlaGlyGlyAlaLeuSerGlyAlaAlaProAlaG 1506
2417CGCGAGCGCAG.....CCACAGAG..... 2437
1506 IyLeuSerLeuArgGlyArgProAlaValProGlyGlyAlaGlySerVal 1522
2437 2437
1523 LeuValLeuLeuGlyAlaAlaGlyAspGlyLeuAspGlyAspGlyGly 1539
2438GGTGTGGCGCCCTTCGCGCGAGCG..... 2464
1539 yGlyValGlyValGlyValGlyValGlyValGlyValAspGlyAlaProGlyAla 1556
2465 ...CAGGAGGCGGTGAACCGAAGTAGACAGAGC..... 2497
1556 yArgProArgGlyGlyProProArgGlyGlyGlyLeuValGlyGlyGly 1572
2498GACGTGCGCGCTCCGATCGGCGTGC 2523
1573 LeuAlaValLeuValMetValThrThrAlaValProSerGlyGlyGly... 1588
2524 GACCTCACCGGGGATAGCTGACCGCGACGACGACGCGCGCGCTG 2573
1589GlyAlaAlaAlaAlaGlyArgArgAspArgProGlyGlyG 1602
2574 GCGGC.....CTCGCGCGCGTCATGA... 2595
1602 IyGlyGlyTrpGlySerGlyProProProCysArgArgGlyHisArg 1618
2596 TGTGGCTGTTTGGACGTCCGCTACGCT..... 2625
1619 CysTrpLeuCysTrpTrpArgArgGlyProArgProArgArgArgProGly 1635
2626GGCGCGCGCTGA 2638
1635 yLeuThrAspArgValProProArgGlyGlyProSerProArgGlyCysA 1652
2639 GCGGTGACGTCGCGCTGAGT..... 2661
1652 rGlyValAlaArgAlaGlyGlyGlyArgGlyGlyCysGlyGlyGly 1668
2662 ...CGTCCGCGTGCATCC...GACCCTGG.....TGATC 2690
1669 ArgAlaProGlyAlaAlaGlyGlyProGlyLeuCysArgCysGlyCysC 1685
2691 GCGTCGAGAGGTGAGATCAACCGCGCGCGCGCGCGCTCCGCGAG 2740
1685 sArgGlyArgArgProGlyProGlyAlaGlyProGlyProGlyProGly 1702
2741 CAGCCGTG.....CCACGACGAGCTG 2763
1702 sPValValThrValLeuGlyAlaLeuMetLeuSerProThrGlyGlyPro 1718
2764 CGAGGTGATGACCTGA 2781
1719 GlyGlyArgGlyProGly 1724
seq_name: SwissProt_40:CA11_CHICK
seq_documentation_block:
ID CA11_CHICK STANDARD; PRT; 1453 AA.
AC P02457;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Collagen alpha 1(I) chain precursor.
GN COL1A1.
OS Gallus gallus (Chicken).


```

915 AACCGGCGGAGCTGATCGGCGG.....CTTGGCGTACCATCTCCGCTG 872
    ||||| ||| ||| |||||
817 uThrglyAspAlaGlyAlaIAspAlaGlyProProGlyProAla 833
    ||||| ||||| ||||| |||||
871 TTGGTATCTGCGAGCTGCTGGTGTGGTGCACGATCCGGCGATGGC 822
    ||| ||| ||||| ||||| |||||
834 ..GlyProThrglyAlaProGlyProAlaGly**ValGlyAlaProGly 849
    ||| ||| ||||| ||||| |||||
821 CCGGAGGCGCGTGCAGCTTCTCAAGC..... 795
    ||||| ||| ||| |||
850 ProGlyGlyAlaArgGlySerAlaGlyProProGlyAlaThrglyPheP 866
    ||||| ||| ||| |||
795 ..... 795
866 oglyAlaAlaGlyArgValGlyProProGlyProSerGlyAsnIleGlyL 883
    ||||| ||||| ||||| |||||
794 ..ACTGGGCTCGGGCGGAGGCGGGGTGAGCGGACGAGCC 747
    ||||| ||||| ||||| |||||
883 euProGlyProProGlyProAlaGlyLys**Gly...SerLysGlyPro 898
    ||||| ||||| ||||| |||||
746 TGCCGGGGGCGTCCGAGACACTCGGC...CTTGAGAGCGCTCTCTCG 700
    ||||| ||||| ||||| |||||
899 ...ArgGlyLysThrglyProAlaGlyArgProGlyGlyLysProAl 914
    ||||| ||||| ||||| |||||
699 AAGCGGTGCACTACGCCGCGGAGACA...CACCGGACCATGAC... 657
    ||||| ||||| ||||| |||||
914 aclyProProGlyProProGlyLysSerProGlyAlaAspGlyP 931
    ||||| ||||| ||||| |||||
656 ..... CCGGCTGCTGACGAGCGCGCA 633
    ||||| ||||| ||||| |||||
931 rolleGlyAlaProGlyThrglyProGlyProGlyIleAlaGlyLys 947
    ||||| ||||| ||||| |||||
632 GCGCGAGTTCGGCTGGTCCAGACAGCACTGCTATCATGATCACC 583
    ||| ||| ||| ||| |||
948 GlyValAlaGlyLeu.....ProGlyLys 955
    ||| ||| ||| ||| |||
582 GGCATCACTTCGGCGGACAGACACAGGCTCTCTGGGCTCTG 533
    ||| ||| ||| ||| |||
956 .....ArgGlyLysArgGlyPheProGlyLeuPro 966
    ||| ||| ||| ||| |||
532 CTCGCGGAGT...CTTGGC.....GGGCGGCTCGCGGATGCGCA 492
    ||| ||||| ||||| |||||
966 LProSerGlyLysProGlyLysGlyLysProSerGlyAlaSerGlyL 982
    ||||| ||||| ||||| |||||
491 CGAGAGCGCGCTCTCCGGTTCGGAGAGAGCGCTCGTACACCGC 442
    ||||| ||||| ||||| |||||
983 ArgGlyProProGlyProMetGlyProProGlyLeuAlaGlyProPro 999
    ||||| ||||| ||||| |||||
441 CCGTGCCTACAGCTGTGGAGTTCGTCGACAGAGTACCATCGC 392
    ||| ||| ||| ||| |||
999 yGlyAlaGlyArgGlyLysAlaProGlyAlaGlyLysAlaProGlyArg 1016
    ||||| ||||| ||||| |||||
391 GCGCTCGCGCTCGCGCGGAGC..... 369
    ||||| ||||| ||||| |||||
016 spGlyAlaAlaGlyProLysGlyAspArgGlyLysThrglyProAlaGly 1032
    ||||| ||||| ||||| |||||
368 .....GC 367
    ||||| ||||| ||||| |||||
1033 ProProGlyAlaProGlyAlaProGlyAlaProGlyProValGlyProAl 1049
    ||||| ||||| ||||| |||||
366 CGGTGCTGTGATGACATGAGGAGCAAA.....CACCGAGCGC 329
    ||||| ||||| ||||| |||||
1049 aclyLysAsnGlyAspArgGlyLysThrglyProAlaGlyProAlaGly 1066
    ||||| ||||| ||||| |||||
1066 roProGlyProAlaGlyAlaArgGlyProAlaGlyProGlyLysProArg 1082
    ||||| ||||| ||||| |||||
282 GCGCGC..... 277
    ||||| ||||| ||||| |||||
1083 GlyAspLysGlyLysThrglyLysGlyLysAspArgGlyLysGlyLys 1099
    ||||| ||||| ||||| |||||
276 .....GGCTCACTTCGGCGAGCG 258
    ||||| ||||| ||||| |||||

```

```

1099 sArgGlyPheSerGlyLeuGlyLysProGlyProProGlyAlaProG 1116
    ||| ||||| ||||| |||||
257 GCCGCACTACTGATCGGGAGGAGCAGCTGCCAGCTGAGTCGCGCAG 208
    ||||| ||||| ||||| |||||
1116 LysGlyLysProSerGlyAlaSerGlyProAlaGlyProArgGlyPro 1132
    ||||| ||||| ||||| |||||
207 TCATGGCGCTACTGGCAGCAGCTTC.....CCGAG 176
    ||||| ||||| ||||| |||||
1133 ProGlySerAlaGlyAlaAlaGlyLysAspGlyLeuAsnGlyLeuPro 1149
    ||||| ||||| ||||| |||||
175 GCCGCACTGCGCGCTGCCGACGAGTTCGCTGCGCGAGGAGCGC 126
    ||||| ||||| ||||| |||||
1149 yProLleGlyPro.....ProGlyP 1156
    ||||| ||||| ||||| |||||
125 CCAGAGCGCGCGCTCACCGAGCTGCTGCTGCTGCTGCTGCTGCTG 76
    ||||| ||||| ||||| |||||
1156 roArgGlyArgThrglyLysValGlyProValGlyProProGlyPro 1172
    ||||| ||||| ||||| |||||
75 GCCGCGACCGCG...ACGGAGCGCGCGCACC 49
    ||||| ||||| ||||| |||||
1173 GlyProProGlyProProGlyProProSer 1182
    ||||| ||||| ||||| |||||

seq_name: SwissProt_40:CA11_CANFA

seq_documentation_block:
ID CA11_CANFA STANDARD; PRT; 1460 AA.
AC 09XSJ7;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Collagen alpha 1(I) chain precursor.
GN COL1A1
OS Canis familiaris (dog)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_Taxid=9615;
RN [1]
RP SEQUENCE FROM N.A.
RC Tissue=Skin;
RA Campbell B.G., Woollon J.A.M., McLeod J.N., Minor R.R.;
RT "Sequence of normal canine COL1A1 cDNA.";
RL Submitted (MAY-1999) to the EMBL/Genbank/DBJ databases
CC -!- FUNCTION: TYPE I COLLAGEN IS A MEMBER OF GROUP I COLLAGEN
CC (FIBRILLAR FORMING COLLAGEN).
CC -!- SUBUNIT: TRIMERS OF ONE ALPHA 2(I) AND TWO ALPHA 1(I) CHAINS.
CC -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC -!- SIMILARITY: CONTAINS 1 WFC DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, AF153062; A034619.1; -
DR InterPro; IPR000087; Collagen.
DR InterPro; IPR000885; Fib_collagen_C.
DR InterPro; IPR001007; WFC.
DR Pfam; PF01410; COLF1; 1.
DR Pfam; PF01391; collagen; 18.
DR ProDom; PD002078; Fib_collagen_C; 1.
DR SMART; SM00038; COLF1; 1.
DR SMART; SM00214; WFC; 1.
DR PROSITE; PS01208; WFC; 1.
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Glycoprotein; Collagen; Signal.
FT SIGNAL 1 22 BY SIMILARITY.
FT PROPEP 23 157 AMINO-TERMINAL PROPEPTIDE.
FT CHAIN 158 1214 COLLAGEN ALPHA 1(I) CHAIN.

```

FT PROPER 1215 1460 CARBOXYL-TERMINAL PROPEPTIDE.
 FT DOMAIN 34 92 VMFC.
 FT DOMAIN 158 174 NONHELICAL REGION (N-TERMINAL).
 FT DOMAIN 175 1188 TRIPLE-HELICAL REGION.
 FT DOMAIN 1189 1214 NONHELICAL REGION (C-TERMINAL).
 FT SITE 741 743 CELL ATTACHMENT SITE (POTENTIAL).
 FT SITE 1089 1091 CELL ATTACHMENT SITE (POTENTIAL).
 FT CARBOHYD 1361 1361 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SO SEQUENCE 1460 AA: 138762 MW: 58E3674D2B570697 CRC64:

alignment_scores:

Quality: 464.50 Length: 1115
 Ratio: 0.940 Gaps: 61
 Percent Similarity: 44.305 Percent Identity: 26.547

alignment_block:

09-673-254-1/rev x CAIL_CANFA ..

Align seg 1/1 to: CAIL_CANFA from: 1 to: 1460

2780 CGAGT CAATCACTCGCAGGCTTCGTGGCAGGC 2743
 174 ProGlyProMetGlyProSerGlyProArgGlyLeuProGlyProGly 190
 2742 TCCTGCGGA GCGGCTGCGGCGGCGGCGTGATC 2708
 190 yAlaProGlyProGlnGlyPheGlnGlyProProGlyGluProGlyGluP 207
 2707 CT CACCTGCTCCGACGCGTACACCGGCGG..... 2676
 207 roGlyAlaSerGlyProMetGlyProArgGlyProProGlyProGly 223
 2675 GATCGACCGCGGACGACCTCAACGGG 2650
 224 LysAsnGlyAspAspGlyLysLysLysProGlyArgProGlyGluArg 240
 2649 A CGTCACCGCTACAGCGCGG..... CCAGCGTACGCGCAG 2612
 240 gGlyProProGlyProGlnGlyAlaArgGlyLeuProGlyThrAlaGlyL 257
 2611 TCCAACAGCCACATCATGACCGCGG..... GGAGCGCG 2574
 257 eu...ProGlyMetLysGlyHisArgGlyPheSerGlyLeuAspGlyAla 272
 2573 CAGGCGCTGCGGCGGAGCTGCG..... 2550
 273 LysGlyAspAlaGlyProAlaGlyProLysGlyGluProGlySerProG 289
 2549 GTCAGCTATCACCGCGTG..... A 2529
 289 yGluAsnGlyAlaProGlyGlnMetGlyProArgGlyLeuProGlyGluA 306
 2528 GGTCCGACCCCGCATCGAGCGGCGGCTGCTCTACTTCGGT 2479
 306 rgGlyArgProGlyAlaProGlyProAlaGlyAlaArgGlyLysAsnGly 322
 2478 TCACACCCCTTCTGCGCTCGGCGGCAAGGCGCGCACCTCTGCTGG 2429
 323 AlaThrGlyAlaAlaGlyPro.....ProGlyProThr..... 333
 2428 CTGGCTTCGCGCGCGGAGGTGACACGCGGCGGCTACTACAGCA 2379
 334 GlyProAlaGlyProProGlyPheProGlyAlaValGlyAlaLys 349
 2378 CCGGCGGCTGTCGCCGAGCGGCGCGACCGCGCGCGGCTGCGG 2329
 349 LysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 365
 2328 CGAAGCTTCGAGGCGGCGGCGGCGG..... CCGTCGCG 2294
 366 ArgGlyLysLysLysLysLysLysLysLysLysLysLysLysLysLys 382

2293 GACACCGCGCAGTACGCG..... GCGGCGCGCGCGCGCGCGCATGTC 2250
 382 yAsnProGlyAlaAspGlyLysProGlyAlaLysGlyAlaAsnGlyAla 398
 2249 GTCTCATCCCGAGATGCTCTCTATCCGCGGACGACGCTGCTGG 2200
 399 ProGlyLysLysLysLysLysLysLysLysLysLysLysLysLysLys 411
 2199 CGGATTCATCGAAGGAGGATTCGTGACAGGCTGCGCGCGCGCGCC 2150
 412 GlyProSerGlyP 416
 2149 CGCATGTTGACCTGCTTCGCCGATGTCGCCGCTGCGCGCATCTCT 2100
 416 roGlnGly.....ProSerGlyProProGlyProLysGly 427
 2099 ACACGCGCTGTCGCTGAGCGGCGGACCGGTGTACGACCGCGGCGC 2050
 428 AsnSerGlyLysProGlyAlaProGlyAsnLysGlyAspThrGlyAlaLys 444
 2049 GCGGATACACGACCTTCAGCTCCGAGGAGGAGCTGTCGCGCGGCTGC 2000
 444 sGlyLys.....ProGlyProThrGlyLysGlnGlyProProGlyP 458
 1999 CCGCC.....ACGATGACACACCCCTTACAC 1971
 458 roAlaGlyLysLysLysLysLysLysLysLysLysLysLysLysLysLys 474
 1970 GGC.....CGTACGCGCGCGGAGAC..... 1950
 475 GlyLeuProGlyProProGlyLysLysLysLysLysLysLysLysLys 491
 1949ACCAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1907
 491 eProGlyAlaAspGlyAlaLysLysLysLysLysLysLysLysLysLys 508
 1906 CGACGAGCGCGAGCGACTACACAGCGGCTCGAGAGCGGCGGCGGAG 1857
 508 LysSerProGlyProAlaGlyProLysGlySerProGlyGluAlaGlyLys 524
 1856 TCATCATGACTCCGATGACGCTCTGCGCGCGGCGGCGGCGGCTTC 1807
 525 ProGly.....GluAlaGlyLeuProGlyAlaLysGlyLysLeuThrLys 539
 1806 GCGGACCGCGCGCGCGCGGCTGCGGCGGCGGCGGCGGCGGCGGCGG 1757
 539 rProGlySerProGlyProAspGlyLysThrGly..... 550
 1756 GGGCGGAGAGTGACGCGGCTGCGGCGGCTGCGGCTGCGGCTGAGCTG 1707
 551ProProGlyProAlaGlyLysAlaGly 559
 1706 TGACGAGCGGATCGGCGCGCGGCGGACTTACCGCGGAGCGGCGGCG 1657
 560 ArgProGlyProProGlyProProGlyAla.....ArgLysLysLys 573
 1656 CTGCTTCGCGCGGCGGCGGAGGAGGCGGCTACCGACCGCGGCTGCGCA 1607
 573 agLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 590
 1606 CATCGGTGACACACCGGCTGCGGCGGCGGAGGCTGCGGCTGCGGCGG 1557
 590 yAlaLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 606
 1556CGTACGCGCGGCGGCGGCGGCTGCTTCGCGGCGGCGG 1519
 607 GlyLysAspGlyLysLysLysLysLysLysLysLysLysLysLysLys 622
 1518 GACTGAGCGGCGGCGG.....TCGCGGCGGCGGCGGCGGCGGCGG 1484
 622 yProAlaGlyLysLysLysLysLysLysLysLysLysLysLysLysLys 639
 1483 CCGGCGCTCTCGAGAGAGCGCGCGGCGGCGGCGGCGGCGGCGG 1434

```
639 lnglyleuprogllyproalaglyproprogllylunalaaglylserproglly 655
1433 CGG.....ACGTGCTGACCCCGGAGGGGCGGCTTCTGCTGGTGCAG 1390
656 glunlgllyalprogllyasplunlgllyalaprogllyproserlyalaa 672
1389 CTCGGGGAGGATACCCCGGCGGGCGGGGTG..... 1356
672 gglglunargllypneprogllylunargllyalaglunlgllyproproglly 699
1355 .....CCTGTAGAGCGGAGGCGCCCGGGGTGCGTGCAGCCGTT 1314
689 rolaaglyproarglyalaaanglyalaprogllyasnasp..... 702
1313 CGGTGTCCCATGATGACCATGAGCGCAAGCCCGAGTGCAGACGCT 1264
703 glyalalaglyalaspalaglyalaprogllyalaprogllyserlunlgl 719
1263 TCCGGGA.....GGCGGGCCCGGTGTCGAGGTGAACGC..... 1230
719 aoprogllyleuglnlglmetprogllylunargllyalalaglyleuprog 736
1229 ..CCCCGGGCGGACCCGCTGGTATCACCAGTACGCCCTGCGCCG 1182
736 lyprolysgly.....Asparglyalaspalaglypro 746
1181 CGAGGTGCTGGCGATCCCGGTGCTGAAGACCCGA..... 1143
747 lysglyalalaspalyserprogllylaspalyalarglyleuthrcl 763
1142 .....CTGCGCCCGCGCGCTGGCGGGGGT...GAGCAGC 1109
763 yprolleglyproprogllypro.....Alaglyalaprogllyaspl 777
1108 GGCTGTGACATCCCGTTCGGAGCTGCTCCGTTACGCTATGCC... 1062
777 ysglylunalaaglyproserglyproalaglyprothrlyalalargly 793
1061 .....CGTGGACGGGAGGCCACGCGGCGCTGCG..... 1032
794 Alaprogllyalasparglylunprogllyproprogllyproalaglyheal 810
1031 .....CCGATCCACGACCTGCTTCACCCGCGC 1001
810 aglyproprogllyalaspalylnprogllyalalysglylunproglly 827
1000 CGGTGGCGGAGGAGGATGCGATGCGGGATGCGCGGCGGCTGCT 951
827 spalaglyAlalysglyaspAlaglyproprogllyproalagly.... 841
950 CACCGAATCGCGACGCTCGCGCGGTGGGCAACCGCGCGAGCTGA 901
842 ..Prothrcllyproprogllyprolellyasnvalgllyalaprogllypro 857
900 TGGGGGCTCGCGTACCACTCCCGCTGTGTCATCTGCGAGCTGTC 851
858 lysglyalAlaArgglyseralaglyproprogllyalathrglypneprog 874
850 GGTGTCCGGTACCGATCCGGCGATCGCGGAGCGGCTGACGGTCT 801
874 lyAlalalaglylunarglylunprogllyproserglyasnalaaglypro 890
800 CAGGCACTCGGCTCGCGCGCGCAGAGCGGGGGGTGACGG..... 756
891 Proglly.....Proprogllyproalaglylunlgllylgllylgllyal 905
755 .....CACGACCTCGCGGGGGGTGCTCGGACACCTCGGC...C 719
905 aarglylunlthrglylunproAla...glylunprogllylunlgllylun 921
718 CTGAGAGAGCTGCTCTCGAAGCGGTGACCTGAGCCGCGGCGAAGACAC 669
||||| |||||:||||| :|||
639 lnglyleuprogllyproalaglyproprogllylunalaaglylserproglly 655
1433 CGG.....ACGTGCTGACCCCGGAGGGGCGGCTTCTGCTGGTGCAG 1390
656 glunlgllyalprogllyasplunlgllyalaprogllyproserlyalaa 672
1389 CTCGGGGAGGATACCCCGGCGGGCGGGGTG..... 1356
672 gglglunargllypneprogllylunargllyalaglunlgllyproproglly 699
1355 .....CCTGTAGAGCGGAGGCGCCCGGGGTGCGTGCAGCCGTT 1314
689 rolaaglyproarglyalaaanglyalaprogllyasnasp..... 702
1313 CGGTGTCCCATGATGACCATGAGCGCAAGCCCGAGTGCAGACGCT 1264
703 glyalalaglyalaspalaglyalaprogllyalaprogllyserlunlgl 719
1263 TCCGGGA.....GGCGGGCCCGGTGTCGAGGTGAACGC..... 1230
719 aoprogllyleuglnlglmetprogllylunargllyalalaglyleuprog 736
1229 ..CCCCGGGCGGACCCGCTGGTATCACCAGTACGCCCTGCGCCG 1182
736 lyprolysgly.....Asparglyalaspalaglypro 746
1181 CGAGGTGCTGGCGATCCCGGTGCTGAAGACCCGA..... 1143
747 lysglyalalaspalyserprogllylaspalyalarglyleuthrcl 763
1142 .....CTGCGCCCGCGCGCTGGCGGGGGT...GAGCAGC 1109
763 yprolleglyproprogllypro.....Alaglyalaprogllyaspl 777
1108 GGCTGTGACATCCCGTTCGGAGCTGCTCCGTTACGCTATGCC... 1062
777 ysglylunalaaglyproserglyproalaglyprothrlyalalargly 793
1061 .....CGTGGACGGGAGGCCACGCGGCGCTGCG..... 1032
794 Alaprogllyalasparglylunprogllyproprogllyproalaglyheal 810
1031 .....CCGATCCACGACCTGCTTCACCCGCGC 1001
810 aglyproprogllyalaspalylnprogllyalalysglylunproglly 827
1000 CGGTGGCGGAGGAGGATGCGATGCGGGATGCGCGGCGGCTGCT 951
827 spalaglyAlalysglyaspAlaglyproprogllyproalagly.... 841
950 CACCGAATCGCGACGCTCGCGCGGTGGGCAACCGCGCGAGCTGA 901
842 ..Prothrcllyproprogllyprolellyasnvalgllyalaprogllypro 857
900 TGGGGGCTCGCGTACCACTCCCGCTGTGTCATCTGCGAGCTGTC 851
858 lysglyalAlaArgglyseralaglyproprogllyalathrglypneprog 874
850 GGTGTCCGGTACCGATCCGGCGATCGCGGAGCGGCTGACGGTCT 801
874 lyAlalalaglylunarglylunprogllyproserglyasnalaaglypro 890
800 CAGGCACTCGGCTCGCGCGCGCAGAGCGGGGGGTGACGG..... 756
891 Proglly.....Proprogllyproalaglylunlgllylgllylgllyal 905
755 .....CACGACCTCGCGGGGGGTGCTCGGACACCTCGGC...C 719
905 aarglylunlthrglylunproAla...glylunprogllylunlgllylun 921
718 CTGAGAGAGCTGCTCTCGAAGCGGTGACCTGAGCCGCGGCGAAGACAC 669
||||| |||||:||||| :|||
921 rogllyproprogllyproalaglylunlgllylserprogllyalalaspaly 937
668 CCGGACCATGACCCGCTGCTGACGAGCGCGCAGCGGAGCTGCGCT 619
938 Proalaglyalapro..... 942
618 CGGTCTCCGACGACCAAGCTGCTACATGATCAACCGGCTCATCTGCC 569
943 ..glylunlthrglylunlgllylunlgllylunlgllylunlgllyl 959
568 GGCCA.....CGACACACCGGCTCTCTGCTGCTGCTGCTGCTGCT 531
959 eutrogllylunarglylunlgllylunlgllylunlgllylunlgllyl 975
530 ..... 507
976 glylunprogllylunlgllylunlgllylunlgllylunlgllylunlgl 992
522 TCCTGGC.....GGCGCGCT..... 507
992 ooprogllylunlgllylunlgllylunlgllylunlgllylunlgllyl 1009
506 ..CGCGCGGATGCGGACGAGGAGCGGCTGCTGCTGCTGCTGCTGCTGCT 462
1009 lyarglylunlgllylunlgllylunlgllylunlgllylunlgllyl 1025
461 .....GGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 442
1026 Progllyprolysglylunlgllylunlgllylunlgllylunlgllylunlgl 1042
441 C.....GGTGGCTGACGTTGTGAGGTGCTGCTGCTGCTGCTGCTGCTGCT 398
1042 yalaprogllyalA..... 1046
397 ATCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 348
1047 .....Progllyalaprogllylunlgllylunlgllylunlgllylunlgl 1059
347 GGGACCAACACGAGCGGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 298
1060 glylunlgllylunlgllylunlgllylunlgllylunlgllylunlgllyl 1071
297 CGGACCGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 248
1071 aglyprolellylunlgllylunlgllylunlgllylunlgllylunlgllyl 1079
247 TGCATGCGGAGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 198
1080 .....Argglylunlgllylunlgllylunlgllylunlgllylunlgllyl 1094
197 ACTGCGCAGAGTTCCTGCGGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 151
1095 thrlylunlgllylunlgllylunlgllylunlgllylunlgllylunlgllyl 1111
150 ACTTGGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 122
1111 yleuglnlgllylunlgllylunlgllylunlgllylunlgllylunlgllyl 1128
121 AGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 82
1128 erglylunlgllylunlgllylunlgllylunlgllylunlgllylunlgllyl 1144
81 CCGCGGCGGCGGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 53
1145 serprogllylunlgllylunlgllylunlgllylunlgllylunlgllylunlgl 1161
52 CACCGCGGACGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 22
1161 oglylunlgllylunlgllylunlgllylunlgllylunlgllylunlgllyl 1173
seq_name: SwissProt_40:CA11_HUMAN
seq_documentation_block:
```


ID CALL_HUMAN STANDARD; PRT; 1464 AA.
AC P02452; Q15176; Q14037;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-MAR-1989 (Rel. 10, Last annotation update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Collagen alpha 1(I) chain precursor.
GN COL1A1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE OF 1-472 FROM N.A.
RX MEDLINE=89025644; PubMed=3178743;
RT Tromp G., Kivianiemi H., Stacey A., Shikata H., Baldwin C.T.,
Jansen R., Prockop D.J.;
RL "Structure of a full-length cDNA clone for the prepro alpha 1(I)
chain of human type I procollagen";
RN Biochem. J. 253:919-922(1988).
RN [2]
RP SEQUENCE OF 1-181 FROM N.A.
RX MEDLINE=84270697; PubMed=6462220;
RT Chu M.-L., de Wet W.J., Bernard M.P., Ding J.-F., Morabito M.,
Myers J., Williams C., Ramirez F.;
RL "Human pro alpha 1(I) collagen gene structure reveals evolutionary
conservation of a pattern of introns and exons";
RN Nature 310:337-340(1984).
RN [3]
RP SEQUENCE OF 162-301.
RX TISSUE=SKIN;
RC MEDLINE=71038625; PubMed=5529814;
RA Click E.M., Bornstein P.;
RL "Isolation and characterization of the cyanogen bromide peptides from
the alpha 1 and alpha 2 chains of human skin collagen";
RN Biochemistry 9:4696-4706(1970).
RN [4]
RP SEQUENCE OF 263-268.
RX TISSUE=SKIN;
RC MEDLINE=71001508; PubMed=4319110;
RA Morgan P.H., Jacobs H.G., Segrest J.P., Cunningham L.W.;
RL "A comparative study of glycopeptides derived from selected
vertebrate collagens. A possible role of the carbohydrate in fibril
formation";
RN J. Biol. Chem. 245:5042-5048(1970).
RN [5]
RP SEQUENCE OF 425-1464 FROM N.A.
RX MEDLINE=84080385; PubMed=6689127;
RT Bernard M.P., Chu M.-L., Myers J.C., Ramirez F., Eikenberry E.F.,
Prockop D.J.;
RL "Nucleotide sequences of complementary deoxyribonucleic acids for the
pro alpha 1 chain of human type I procollagen. Statistical evaluation
of structures that are conserved during evolution";
RN Biochemistry 22:5213-5223(1983).
RN [6]
RP SEQUENCE OF 1229-1454 FROM N.A.
RX TISSUE=BONE;
RC MEDLINE=88124208; PubMed=3340531;
RA Meekelae J.K., Raassina M., Virta A., Vuorio E.;
RL "Human pro alpha 1(I) collagen: cDNA sequence for the C-propeptide
domain";
RN Nucleic Acids Res. 16:349-349(1988).
RN [7]
RP SEQUENCE OF 1-34 FROM N.A.
RX MEDLINE=88097389; PubMed=3480516;
RT Bornstein P., McKay J., Morishima J.K., Devarayalu S., Gelinas R.E.;
RL "Regulatory elements in the first intron contribute to
transcriptional control of the human alpha 1(I) collagen gene";
RN Proc. Natl. Acad. Sci. U.S.A. 84:8869-8873(1987).
RN [8]
RP SEQUENCE OF 1-34 FROM N.A.
RX MEDLINE=85130970; PubMed=2857713;
RT Chu M.-L., de Wet W.J., Bernard M.P., Ramirez F.;
RL "Fine structural analysis of the human pro-alpha 1 (I) collagen gene."

RT Promoter structure, AluI repeats, and polymorphic transcripts.";
RN J. Biol. Chem. 260:2315-2320(1985).
RN [9]
RP SEQUENCE OF 1-44 FROM N.A.
RX MEDLINE=88033098; PubMed=2822714;
RT Rossouw C.M.S., Vergeer W.P., du Plooy S.J., Bernard M.P., Ramirez F.,
de Wet W.J.;
RL "DNA sequences in the first intron of the human pro-alpha 1(I)
collagen gene enhance transcription";
RN J. Biol. Chem. 262:15151-15157(1987).
RN [10]
RP REVIEW ON VARIANTS.
RX MEDLINE=91184577; PubMed=2010058;
RT Kivianiemi H., Tromp G., Prockop D.J.;
RL "Mutations in collagen genes: causes of rare and some common diseases
in humans";
RN FASEB J. 5:2052-2060(1991).
RN [11]
RP REVIEW ON VARIANTS.
RX MEDLINE=97255959; PubMed=9101290;
RT Kivianiemi H., Tromp G., Prockop D.J.;
RL "Mutations in fibrillar collagens (types I, II, III, and XI), fibril-
associated collagen (type IX), and network-forming collagen (type X)
cause a spectrum of diseases of bone, cartilage, and blood vessels";
RN Hum. Mutat. 9:300-315(1997).
RN [12]
RP REVIEW ON VARIANTS.
RX MEDLINE=91374476; PubMed=1895312;
RT Byers P.H., Wallis G.A., Willing M.C.;
RL "Osteogenesis imperfecta: translation of mutation to phenotype";
RN J. Med. Genet. 28:433-442(1991).
RN [13]
RP REVIEW ON VARIANTS.
RX MEDLINE=97169389; PubMed=9016532;
RT Dalgleish R.;
RL "The human type I collagen mutation database";
RN Nucleic Acids Res. 25:181-187(1997).
RN [14]
RP VARIANT OI-II CYS-1166.
RX MEDLINE=86287390; PubMed=3016737;
RT Conn D.H., Byers P.H., Steinmann B., Gelinas R.E.;
RL "Lethal osteogenesis imperfecta resulting from a single nucleotide
change in one human pro alpha 1(I) collagen allele";
RN Proc. Natl. Acad. Sci. U.S.A. 83:6045-6047(1986).
RN [15]
RP VARIANT OI-II ARG-569.
RX MEDLINE=87222295; PubMed=3108247;
RT Bateman J.F., Chan D., Walkers I.D., Rogers J.G., Cole W.G.;
RL "Lethal perinatal osteogenesis imperfecta due to the substitution of
arginine for glycine at residue 391 of the alpha 1(I) chain of type I
collagen";
RN J. Biol. Chem. 262:7021-7027(1987).
RN [16]
RP VARIANT OI-II CYS-926.
RX MEDLINE=88033031; PubMed=3667599;
RT Vogel B.E., Minor R.R., Freund M., Prockop D.J.;
RL "A point mutation in a type I procollagen gene converts glycine 748
of the alpha 1 chain to cysteine and destabilizes the triple helix in
a lethal variant of osteogenesis imperfecta";
RN J. Biol. Chem. 262:14737-14744(1987).
RN [17]
RP VARIANT OI-II ARG-842.
RX MEDLINE=88298828; PubMed=3403550;
RT Bateman J.F., Lemande S.R., Dahl H.H., Chan D., Cole W.G.;
RL "Substitution of arginine for glycine 664 in the collagen alpha 1(I)
chain in lethal perinatal osteogenesis imperfecta. Demonstration of
the peptide defect by in vitro expression of the mutant cDNA";
RN J. Biol. Chem. 263:11627-11630(1988).
RN [18]
RP VARIANT OI CYS-1195.
RX MEDLINE=89218628; PubMed=3244312;
RT Labhard M.E., Wirtz M.K., Pope F.M., Nicholls A.C., Hollister D.W.;
RL "A cysteine for glycine substitution at position 1017 in an alpha

RT 1(I) chain of type I collagen in a patient with mild dominantly
 RT inherited osteogenesis imperfecta.";
 RL Mol. Biol. Med. 5:197-207(1988).
 RN [19]
 RP VARIANT OI-II VAL-434.
 RX MEDLINE-89255493; PubMed-2470760;
 RA Paterson E., Smiley E., Bonadio J.;
 RT "RNA sequence analysis of a perinatal lethal osteogenesis imperfecta
 mutation.";
 RL J. Biol. Chem. 264:11083-11087(1989).
 RN [20]
 RP VARIANT OI-IV SER-1010.
 RX MEDLINE-89305591; PubMed-2745420;
 RA Martin J.C., Grange D.K., Gottesman G.S., Lewis M.B., Koepf D.A.;
 RT "Osteogenesis imperfecta type IV. Detection of a point mutation in
 one alpha 1(I) collagen allele (COL1A1) by RNA/RNA hybrid analysis.";
 RL J. Biol. Chem. 264:11893-11900(1989).
 RN [21]
 RP VARIANTS OI-II ALA-1106; VAL-1151; ARG-1154 AND VAL-1184.
 RX MEDLINE-89380165; PubMed-2777764;
 RA Lamande S.R., Dahl H.-H.M., Cole W.G., Bateman J.F.;
 RT "Characterization of point mutations in the collagen COL1A1 and
 COL2A1 genes causing lethal perinatal osteogenesis imperfecta.";
 RL J. Biol. Chem. 264:15809-15812(1989).
 RN [22]
 RP VARIANT OI SER-1022.
 RX MEDLINE-90062068; PubMed-2511192;
 RA Pack M., Constantinou C.D., Kalia K., Nielsen K.B., Prockop D.J.;
 RT "Substitution of serine for alpha 1(I)-glycine 844 in a severe
 variant of osteogenesis imperfecta minimally destabilizes the triple
 helix of type I procollagen. The effects of glycine substitutions on
 thermal stability are either position of amino acid specific.";
 RL J. Biol. Chem. 264:19694-19699(1989).
 RN [23]
 RP VARIANT OI-II CYS-1082.
 RX MEDLINE-89109573; PubMed-2913053;
 RA Constantinou C.D., Nielsen K.B., Prockop D.J.;
 RT "A lethal variant of osteogenesis imperfecta has a single base
 mutation that substitutes cysteine for glycine 904 of the alpha 1(I)
 chain of type I procollagen. The asymptomatic mother has an
 unidentified mutation producing an overmodified and unstable type I
 procollagen.";
 RL J. Clin. Invest. 83:574-584(1989).
 RN [24]
 RP VARIANTS OI CYS-272; CYS-704 AND CYS-896.
 RX MEDLINE-90009313; PubMed-2794057;
 RA Starman B.J., Eyre D., Charbonneau H., Harrylock M., Weiss M.A.,
 RA Weiss L., Graham J.M., Byers P.H.;
 RT "Osteogenesis imperfecta. The position of substitution for glycine by
 cysteine in the triple helical domain of the pro alpha 1(I) chains of
 type I collagen determines the clinical phenotype.";
 RL J. Clin. Invest. 84:1206-1214(1989).
 RN [25]
 RP VARIANT OI-II CYS-422.

alignment_scores:
 Quality: 462.00 Length: 1122
 Ratio: 0.937 Gaps: 62
 Percent Similarity: 43.939 Percent Identity: 26.560

alignment_block:
 US-09-673-254-1/rev x CA11_HUMAN ..
 Align seg 1/1 to: CA11_HUMAN from: 1 to: 1464

2780 CCAGGT.....CAATCACTCGCAGGCTTCTCTGCTGGCAGCGC 2743
 |||||||| |||||
 178 ProglyProMetGlyProSerGlyProArgGlyLeuProGlyProGly 194
 2742 TGTCTCGGGA.....GCGGCTCGCGGCGGCGCGCGTATC 2708
 :||| ||||||
 194 yAlaProGlyProGlnGlyPheGlnGlyProProGlyGlnProGlyGlnP 211

2707 CT.....CACCTGTCGCGACGCGCTACACCGACGCGC..... 2676
 ||||| ||||| |||||
 211 roGlyAlaSerGlyProMetGlyProArgGlyProArgGlyProProGly 227
 2675GATCGACCGCGGACGACGACTCAACGCGC 2650
 228 LysAsnGlyAspAspGlyGlnAlaGlyLysProGlyArgProGlyGln 244
 2649 A.....CGGTACCGCTACAGCGCGG.....CCAGCGTACGCGACG 2612
 |||| ||||| |||||
 244 gelyProProGlyProGlnGlyAlaArgGlyLeuProGlyTyrAlaGly 261
 2611 TCCAACAGCGCCACATCATGACCGCGC.....GGAGCGCGC 2574
 :||| ||||| |||||
 261 eu.....ProGlyMetLysGlyHisArgGlyPheSerGlyLeuAspGlyAla 276
 2573 CAGCGCGTGGCGGACGCTGTGC..... 2550
 :||| ||||| |||||
 277 LysGlyAspAlaGlyProAlaGlyProLysGlyGlnProGlySerProGly 293
 2549GTCACCTATCACCGCGGTG.....A 2529
 :||| ||||| |||||
 293 yGlnAsnGlyAlaProGlyGlnMetGlyProArgGlyLeuProGlyGln 310
 2528 GGTCCGACACCGCCATCGACGCGGACGCGCTGCTGCTACTTCGGT 2479
 ||||| ||||| |||||
 310 rGlyLysProGlyAlaProGlyAlaProGlyAlaArgGlyLysAsnGly 326
 2478 TCACCCCTTCTCTGCGCTCGCGCGGAGGCGCGACACCTCGTGG 2429
 :||| ||||| |||||
 327 AlaTrnGlyAlaAlaGlyPro.....ProGlyProThr..... 337
 2428 CTGGCGTGGCGCGCGGAGGAGTTCACACGCGCGGCTACTACAGCA 2379
 ||||| ||||| |||||
 338GlyProAlaGlyProProGlyPheProGlyAlaValGlyAlaLysG 353
 2378 CCGGCGGCTGTCC..... 2365
 :||| ||||| |||||
 353 LysLysAlaGlyProGlnGlyProArgGlySerGlnGlyProGlnGlyAl 369
 2364 CGGTAGCGGCGCGGACCGCGGACCGCGCTCGCGGACGACTGGGAG 2315
 ||||| ||||| |||||
 370 ArgGlyGlnProGlyProProGlyProAlaGly..... 380
 2314 GCGGCGGCGCGCGCGGTGGCGACACCGCGCACTGACGCG.....GGCG 2271
 ||||| ||||| |||||
 381AlaAlaGlyProAlaGlyAsnProGlyAlaAspGlyGlnProGly 395
 2270 GCCCGCGCGCGCGCATGTCCTTCATCCGCGAGATGCTCATCC 2221
 ||||| ||||| |||||
 396 AlaLysGlyAlaAsnGlyAlaProGlyLeuAlaGlyAlaProGlyPhePr 412
 2220 GCGAGCGCACAGCGCTGTGTGCCGATCCATGGAAGAAGCATGCTGTGC 2171
 |||| ||||| |||||
 412 oGlyAlaArg..... 415
 2170 CAGGTTGCGCGCGCGCGCGCGCATGTTGACCTCGGTTGCCCGATG 2121
 ||||| ||||| |||||
 416GlyProSerGlyProGlnGly.....ProGly 424
 2120 TCGCGCGCTCGCGCGACTTCATACCGCGGCTTGTGCGTGCACCGCAC 2071
 :||| ||||| |||||
 425 GlyProProGlyProLysGly.....AsnSerGlyGlnProGly 436
 2070 GTGTTCAGCGACCGCGCGCGCGGCGATACA.....CGAGTTTCA 2030
 :||| ||||| |||||
 437GlyAlaProGlySerLysGlyAspThrGlyAlaLysGlyGlnProG 452
 2029 CTCGACGAGGAGCTGTGCGCGCGGTCGCGCGCGC..... 1995
 :||| ||||| |||||
 452 LysProValGlyAlaGlnGlyProProGlyProAlaGlyGlnGlnGlyLys 468
 1994ACCAGATGCAACCCCTTACACCGCGC..... 1968

469 ArgGLyAlaArgGLyLupProGLyProThrGLyLeuProGLyProProGL 485
1967 ...CGTACGGCGCGGAGAC.....ACCAGACGGCAT 1937
485 yGluThrGLyGLyProGLySerArgGLyPheProGLyAlaAspGLyAla 502
1936 GCGCGCATGTGACCGGTACTGTGCGCACCGGACGCGGACGCGACTA 1887
502 IAGLyProLysGLyProAlaGLyGLyLupArgGLySerProGLyProAlaGLy 518
1886 CCAAGCGGGTCGAGACGGCGGCGGAGTCATCATGACTCCGATGAC 1837
519 ProLysGLySerProGLyGLyLupArgProGLy.....GluAlaGL 533
836 GTCTCGGCGCGCGGATGGCGGCTTCGCCGACCCCGCGGGCGCG 1787
533 yLeuProGLyAlaLysGLyLeuThrGLySerProGLySerProGLyPro 550
1786 GTTCGCGGTCGTGCGCGCAAGGAGTCATGAGGGCGCGGAGTGACGGCG 1737
550 spGLyLysThrGLy..... 554
1736 TCACCGGCTCGGTGCGCTGGTCGACCTGTGACCGACGCGATCGGGCC 1687
555 ...ProProGLyProAlaGLyGLyLupArgProGLyProProGLyPr 570
1686 GCGCGGACTTTCACCGCGGACCTCGGCTCGGCTCGCGCGGACACCGG 1637
570 oProGLyAla.....ArgGLyGluAlaGLyAlaMetGLyPheProG 584
1636 ACTGAAGGCGTCACCGACCGGTCGCAACATCGGTGACACACCGGTGC 1587
584 yProLysGLyAlaAlaGLyLupProGLyLysAlaGLyGLyLupArgGLyAl 600
1586 CCGGACCCGAGGAGCTGGCGTCACCGCG.....CGTACCG 1549
601 ProGLyProProGLyAlaValaGLyProAlaGLyLysAspGLyGluAlaGL 617
1548 CCGGACTGGCGGCTGTTCGCCGCGGACGACTGCGACCGACCG..... 1503
617 yAlaGLyLupPro...ProGLyProAlaGLyProAlaGLyGLyLupArgGLy 633
1502TCGGCGCGCGGTTGAACCTCGGCGGCTCGGCGGACGACG 1464
633 IuGLyLupProAlaGLySerProGLyPheGLyLupProGLyProAla 649
463 AGCCCGCGACACCGCCAGGCGGCGCGG.....ACCTGCTCGAC 1420
650 GLyProProGLyGLyLupArgGLyLysProGLyGLyLupArgGLyAla 666
1419 CCGGACGGGCGGCTTCGCTGCTGCGAATCGCGGAGGGGTACCCCG 1370
666 pLeuGLyAlaProGLyProSerGLyAlaArgGLyGLyLupArgGLyPheProG 683
1369 GCGCGCGCGCGGTG.....CCTGTCGACGG 1344
683 yGLyLupArgGLyAlaGLyLupProProGLyProAlaGLyProArgGLyAla 699
1343 CGAGCGCGCGCGGTGCGGCGGACCGGTCGCTCCGATGATGACCA 1294
700 AsnGLyAlaProGLyAsnAsp.....GLyAlaLysGLyAspAlaGL 713
1293 TGCAGCGCAAGCCGAGTCGACGACGCTTCGGGA.....GGCGGC 1250
713 yAlaProGLyAlaProGLySerGLyLupArgGLyLupArgGLyAla 730
1249 CCGGTCGTGAGGTGAACG.....CCCGCGCGGAGACCCCG 1212
730 roGLyLupArgGLyAlaAlaGLyLupProGLyProLysGLy..... 743
1211 CTGGGTCATACCGATGACGCGCTCGCGGAGGTGTCGGCATCCCG 1162

744AspArgGLyAspAlaGLyProLysGLyAlaAspLysSerPr 757
1161 GGTTCGTGAAGACCCCGA.....CCTCGCC 1136
757 oGLyLysAspGLyAlaArgGLyLeuThrGLyProLysGLyProProGLyP 774
1135 CCGCGCGCGTGGGGGGT.....GGACGAGGTCGTGACATCCCGCTTC 1089
774 ro.....AlaGLyAlaProGLyAspLysGLyLupSerGLyProSer 787
1088 GGAGTCGCTCGGTTCAACGCTCATCGC.....CGTGAGC 1054
788 GLyProAlaGLyProThrGLyAlaArgGLyAlaProGLyAspArgGLyGL 804
1053 GCGAGGCCACCGCGCGCTCGC.....C 1031
804 uProGLyProProGLyProAlaGLyPheAlaGLyProProGLyAlaAspG 821
1030 CCGATCCAGCACCTCGTTCAACCGCGCGCGCTGCGGAGCGAGCGG 981
821 yGLyLupProGLyAlaLysGLyLupProGLyAspAlaGLyAlaLysGLyAs 837
980 TCGCATCGCGCGGATCGCGCGGCTGCTCAACGAACTCGCGGACGCT 931
837 pAlaGLyProProGLyProAlaGLy.....ProAlaGLyProProGLy 851
930 CCGCGCGTCCGGAACCGCGGAGCTATCGCGGCTTCGCTCGAC 881
852 ProLysGLyAsnValaGLyAlaProGLyAlaLysGLyAlaArgGLySerA 868
880 TTCGCGCTGTGCTCATCGGAGCTGCTCGGTCGCGGTCGACGATCC 831
868 IAGLyProProGLyAlaThrGLyPheProGLyAlaAlaGLyAlaGLy 884
830 GCGATGCGCGCGGAGCGGCTGCTCAAGGACCTCGCGCTCGGCG 781
885 ProProGLyProSerGLyAsnAlaGLyProProGLy.....ProProGL 899
780 GCGCGGAGCGGCGGCGGTGACG.....CAGGACCT 746
899 yProAlaGLyLysGLyLupGLyLysGLyProAlaGLyLupArgGLyLupPro 916
745 GCGGCGGCGGTCGAGACCTCGGC.....CTTGAGAGCTGCTCTCGCA 699
916 Ia...GLyArgProGLyGLyLupArgGLyProProGLyProProGLyProAla 931
698 AGCGGTGACATCAGCCCGGCGGAACACACCCGACCATGACCGCGTGC 649
932 GLyGLyLysGLySerProGLyAlaAspGLyProAlaGLyAlaPro..... 946
648 TGTACGAGCGCGCGAGCGGAGTTCGCTCGCTCCGACGACGCTC 599
947GLyThrProGLyProGLy 953
598 GTTCATCATGACCGGCTCATCTTCGCGCGCA.....CGACACAC 555
953 yLysLupArgGLyLupArgGLyAlaGLyLupProGLyLupArgGLyGLy 969
554 CCGCTCTTCCTGCGCTCTCGCGGAGT...CTGCG.....G 514
970 ArgGLyPheProGLyLupProGLyProSerGLyLupProGLyLysGLy 986
513 GCGCGCTCGCGGAGTGGCGAGGAGGACCGCTCCGCGTTCGAGG 464
986 yProSerGLyAlaSerGLyLupArgGLyProProGLyProMetGLyProp 1003
463 GAGGCTGCTTACACCGCGCGGTCCTTACACTGTGAGGTTCG 414
1003 roGLyLupArgGLyPro..... 1009
413 TGCACGAGGTGACCATCGCGCGT..... 387
1010GLyLupSerGLyArgGLyLupArgGLyAlaProGLyAlaGLyLup 1023

```

386 .CCGGCTGCCCGCGAGCGCGCTGCTGTGACATCGAGCGCACCAAC 338
|||||
1023 rProglYAspArglySerProglYAlaLysAspArgly..... 1037
337 ACCGAGCGCGCCATCAGCAGCGCGCGCTTCACCCGCGCGCTCC 288
|||||
1038 .....GluThrnglyProAlaGlyProProglYAlaProglY 1049
287 CTCGTGGCGCGCGCTCACCCTTGGCGAGCGCGCGCTACTGATGGGG 238
1050 AlaProglYAlaPro.....G1 1055
237 AGCAGCTCGCCAGCTGAGTGGCGACGATGATCGCGCTACTGCGCAC 188
|||||
1055 yProValGlyProAlaGlyLysSerGlyAspArglyGluThrnglyPro 1072
187 AGCTTCCCGAGCGCGCGCTGCGCGCTTCACGACGATGGCGGTGTG 138
|||||
1072 laGlyProAlaGlyProValGlyProAlaGlyAlaArglyProAlaGly 1088
137 CCGGAGGCG..... 129
1089 ProGlnnglyProArglyAspLysGlyGluThrnglyGlnGlyAspAr 1105
128 .....GGCCGAGACGGCG 117
1105 gGlylleLysGlyHisArglyPheSerGlyLeuGlnnglyProProglY 1122
116 C....GGTCACCG.....AACTGCGCGCTGCGCTGCGCTATCGCGCG 76
|||||
1122 roProglYSerProglYnglnnglyProSerGlyAlaSerGlyPro... 1137
75 GCGGCGCGCGAGCGCGCGCGCGCGCGCGCG..... 37
1138 ..AlaGlyProArglyProProglYSerAlaGlyAlaProglYLysAs 1153
36 .....GTGGCGCGTCCCGCTGCGCGTGTACCC 12
1153 pGlyLeuAsnglyLeuProglYProIleGlyProProglYProAlaGly 1170
11 GGTGGCGA 4
1170 rGThrgly 1172
seq_name: SwissProt_40:MCU1_HUMAN
seq_documentation_block:
ID MCU1_HUMAN STANDARD; PRT; 1255 AA.
AC P15941; P15942; P15931; P17626; Q14128; Q16442; Q16437; Q9Y4J2;
DT 01-JAN-1990 (Rel. 13, Created)
01-APR-1990 (Rel. 14, Last sequence update)
16-OCT-2001 (Rel. 40, Last annotation update)
Mucin 1 precursor (Polymorphic epithelial mucin) (PEM) (PEMT)
(Epistialin) (Tumor-associated mucin) (Carcinoma associated mucin)
(Tumor-associated epithelial membrane antigen) (EMA) (H23A6) (Peanut-
reactive urinary mucin) (PUM) (Breast carcinoma-associated antigen
DF3).
DE (Human).
GN MCU1.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas;
RX MEDLINE=90368715; PubMed=2394722;
RA Ian M.S., Batra S.K., Qi W.-N., Metzgar R.S., Hollingsworth M.A.;
RT "Cloning and sequencing of a human pancreatic tumor mucin cDNA.";
RL J. Biol. Chem. 265:15294-15299(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=90202794; PubMed=2318825;

```

```

RA Lightberg M.J.L., Vos H.L., Gennissen A.M.C., Hilkens J.;
RT "Epistialin, a carcinoma-associated mucin, is generated by a
RT polymorphic gene encoding splice variants with alternative amino
RT termini.";
RL J. Biol. Chem. 265:5573-5578(1990).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Breast carcinoma;
RX MEDLINE=90368715; PubMed=1697589;
RA Gendler S.J., Lancaster C.A., Taylor-Papadimitriou J., Duhig T.,
RA Peat N., Burchell J., Pemberton L., Lalani E.-N., Wilson D.;
RT "Molecular cloning and expression of human tumor-associated
RT polymorphic epithelial mucin.";
RL J. Biol. Chem. 265:15286-15293(1990).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=91097524; PubMed=2268309;
RA Lancaster C.A., Peat N., Duhig T., Wilson D.,
RA Taylor-Papadimitriou J., Gendler S.J.;
RT "Structure and expression of the human polymorphic epithelial mucin
RT gene: an expressed VNTR unit.";
RL Biochem. Biophys. Res. Commun. 173:1019-1029(1990).
RN [5]
RP SEQUENCE FROM N.A.
RC TISSUE=Breast carcinoma;
RX MEDLINE=90276413; PubMed=2351132;
RA Wreschener D.H., Hareuveni M., Tsarfaty I., Smorodinsky N.,
RA Horov J., Zaretsky J., Kotkes P., Weiss M., Lathe R., Dion A.,
RA Keydar I.;
RT "Human epithelial tumor antigen cDNA sequences. Differential splicing
RT may generate multiple protein forms.";
RL Eur. J. Biochem. 189:463-473(1990).
RN [6]
RP SEQUENCE FROM N.A.
RC TISSUE=Breast carcinoma;
RX MEDLINE=90276414; PubMed=2112460;
RA Hareuveni M., Tsarfaty I., Zaretsky J., Kotkes P., Horev J.,
RA Zrihan S., Weiss M., Green S., Lathe R., Keydar I., Wreschener D.H.;
RT "A transcribed gene, containing a variable number of tandem repeats,
RT codes for a human epithelial tumor antigen. cDNA cloning, expression
RT of the transfected gene and over-expression in breast cancer
RT tissue.";
RL Eur. J. Biochem. 189:475-486(1990).
RN [7]
RP SEQUENCE FROM N.A.
RX MEDLINE=91033045; PubMed=1688329;
RA Tsarfaty I., Hareuveni M., Horev J., Zaretsky J., Weiss M.,
RA Veltsen J.M., Gendler S.J., Lathe R., Keydar I., Wreschener D.H.;
RT "Isolation and characterization of an expressed hypervariable gene
RT coding for a breast-cancer-associated antigen.";
RL Gene 93:313-318(1990).
RN [8]
RP PARTIAL SEQUENCE FROM N.A.
RX MEDLINE=88330762; PubMed=3417635;
RA Gendler S.J., Taylor-Papadimitriou J., Duhig T., Rothbard J.,
RA Burchell J.;
RT "A highly immunogenic region of a human polymorphic epithelial mucin
RT expressed by carcinomas is made up of tandem repeats.";
RL J. Biol. Chem. 263:12820-12823(1988).
RN [9]
RP SEQUENCE OF 1-169 FROM N.A.
RX MEDLINE=90088473; PubMed=2597151;
RA Abe M., Siddiqui J., Kufe D.;
RT "Sequence analysis of the 5' region of the human DF3 breast
RT carcinoma-associated antigen gene.";
RL Biochem. Biophys. Res. Commun. 165:644-649(1989).
RN [10]
RP SEQUENCE OF 1-109 FROM N.A.
RC TISSUE=Thyroid;
RX MEDLINE=96183746; PubMed=8608966;
RA Weiss M., Baruch A., Keydar I., Wreschener D.H.;
RT "Preoperative diagnosis of thyroid papillary carcinoma by reverse
RT transcriptase polymerase chain reaction of the MCU1 gene.";

```



```

756 1sglyValThrSerAlaProAspThrArgProAlaProGlySerThrAla 772
669 CCCCACCATGACCCGCG...TGCTGTACGAGCGCGCCAGCGGAGTTC 623
773 Pro.....ProAlaHisGlyValThrSerAla.....ProAspThr 784
622 GCGTCGGTCTCCGAGACACAGCTGCTACATGATCACC...GGCT 579
784 ArgProAlaPro.....GlySerThrAlaProProAlaHisGly 798
578 CATCTCCCGCGCCACGACACACCGGCTCTCTCGGAGCTTCTGTCG 529
798 alThrSerAlaProAspThrArgProAlaProGlySerThrAlaPro 814
528 CGAGAGTCTCGCGCGCGCTCGCGGGGAGTCCGAGACGAGACCGCTC 479
615 AlaHisGlyValThrSerAla.....ProAspThrArgProAla 827
478 TCCCGGTCTGTGAGAGAGCGCTGCGGTACACACCGCGGTCCCTAC 429
827 aProGlySer.....ThrAlaProProAlaHis 836
428 GTTGTGAGGTTCGCTCCACGAGAGTACATCGCGCGCGCTCGGCTGC 379
836 1sglyValThrSerAlaProAspThrArgProAlaProGlySerThrAla 852
378 CCGCGGAGGCGCGGTGCTGTGTGAGATCGAGGAGGACACACGCGCG 329
853 Pro.....ProAlaHis 856
328 CGCCATACGACGACCGCCGACGCTTCCACCGGACCGCTCCCTCGTGC 279
856 sglyValThrSer.....AlaProAspThrArgProAlaPro...Glys 870
278 CGCGGTCACTTCGCGGAGCGCGCGCTACT..... 247
870 erThrAlaProProAlaHisGlyValThrSerAlaProAspThrArgPro 886
246 GCATCGGGGAGAGCTCGCCACGCTGAGTGGCGGACGATGATCGGCTA 197
887 AlaProGlySerThrAlaPro.....ProAlaHis 896
196 CTGCGACGACAGTTCGCCG.....AGCCCGAGTGGCGGTCCCTAGA 153
896 sglyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaPro 913
152 CGAGTTCGCGGTGTCGCGGAGGCGCCACGAGCGCGCGGTCAACGAC 103
913 rOProAlaHisGlyValThrSerAlaProAspThrArgProAlaPro... 928
102 TGCCCGTCTGCTGCGGTGATCCGCGCGCGGACCGGACGCGCGCC 53
929 .....GlySerThrAlaProProAlaHisGlyValThrSerAlaPro 942
52 CACCGGACGCGCGGTGCGGTCCGCGCG 19
942 cAspAsnArgProAlaLeuGlySerThrAlaPro 953
seq_name: SwissProt_40:CA11_MOUSE
seq_documentation_block:
ID CA11_MOUSE STANDARD: PRT: 1453 AA.
AC P11087: 060635:
DT 01-JUN-1989 (Rel. 11, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Collagen alpha 1(I) chain precursor.
GN COL1A1 OR COLA1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN=FVB/N.
RX MEDLINE=96033240; PubMed=8535610;
RA Li S.W., Khillan J., Prockop D.J.;
RT "The complete cDNA coding sequence for the mouse pro alpha 1(I) chain
RT of type I procollagen."
RL Matrix Biol. 14:593-595(1995).
RN [2]
RP SEQUENCE OF 518-1128 FROM N.A.
RX MEDLINE=86137403; PubMed=3841523;
RA French B.T., Lee W.-H., Maul G.G.;
RT "Nucleotide sequence of a cDNA clone for mouse pro alpha 1(I)
RT collagen protein."
RL Gene 39:311-312(1985).
RN [3]
RP SEQUENCE OF 735-1130 FROM N.A.
RX MEDLINE=83141374; PubMed=6298597;
RA Monson J.M., Friedman J., McCarthy B.J.;
RT "DNA sequence analysis of a mouse pro alpha 1 (I) procollagen gene:
RT evidence for a mouse B1 element within the gene."
RL Mol. Cell. Biol. 2:1362-1371(1982).
RN [4]
RP SEQUENCE OF 735-878 AND 1005-1058 FROM N.A.
RX MEDLINE=83157109; PubMed=6219867;
RA Monson J.M., McCarthy B.J.;
RT "Identification of a Balb/c mouse pro alpha 1(I) procollagen gene:
RT evidence for insertions or deletions in gene coding sequences."
RL DNA 1:59-69(1981).
RN [5]
RP SEQUENCE OF 1442-1453 FROM N.A.
RX MEDLINE=88124276; PubMed=3340560;
RA Mooslehner K., Harbers K.;
RT "Two mRNAs of mouse pro alpha 1(I) collagen gene differ in the size
RT of the 3'-untranslated region."
RL Nucleic Acids Res. 16:773-773(1988).
CC [1] FUNCTION: TYPE I COLLAGEN IS A MEMBER OF GROUP I COLLAGEN
CC (FIBRILLAR FORMING COLLAGEN).
CC [2] SUBUNIT: TRIMERS OF ONE ALPHA 2(I) AND TWO ALPHA 1(I) CHAINS.
CC [3] TISSUE SPECIFICITY: FORMS THE FIBRILS OF TENDON, LIGAMENTS AND
CC BONES. IN BONES THE FIBRILS ARE MINERALIZED WITH CALCIUM
CC HYDROXYAPATITE.
CC [4] PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC [5] SIMILARITY: CONTAINS 1 VWF-C DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U08020; AAA88912.1; -
DR EMBL: X15896; CAA33904.1; -
DR EMBL: M14423; AAA37333.1; -
DR EMBL: M17491; AAA37334.1; -
DR EMBL: X06753; CAA29227.1; -
DR EMBL: K03036; AAA37332.1; -
DR EMBL: K03029; AAA37332.1; JOINED.
DR EMBL: K03030; AAA37332.1; JOINED.
DR EMBL: K03031; AAA37332.1; JOINED.
DR EMBL: K03032; AAA37332.1; JOINED.
DR EMBL: K03033; AAA37332.1; JOINED.
DR EMBL: K03034; AAA37332.1; JOINED.
DR EMBL: K03035; AAA37332.1; JOINED.
DR PIR: A23982; A23982.
DR MGD: MGI:88467; Colla1.
DR InterPro: IPR000087; Collagen.
DR InterPro: IPR000885; Fib-collagen_C.
DR InterPro: IPR001007; VWF_C.
DR Pfam: PF01410; COLF1; 1.
DR Pfam: PF01391; Collagen; 18.

```

DR ProdDom: PD002078; Fib.collagen_C; 1.
 DR SMART: SM00038; COLFI; 1.
 DR SMART: SM00214; VMC; 1.
 DR PROSITE: PS01208; VMEC; 1.
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
 Glycoprotein; Collagen; Signal.
 FT SIGNAL 1 22
 FT PROPEP 23 151 AMINO-TERMINAL PROPEPTIDE.
 FT CHAIN 152 1207 COLLAGEN ALPHA 1(I) CHAIN.
 FT PROPEP 1208 1453 CARBOXYL-TERMINAL PROPEPTIDE.
 FT DOMAIN 29 87 VMEC.
 FT DOMAIN 152 167 NONHELICAL REGION (N-TERMINAL).
 FT DOMAIN 168 1181 TRIPLE-HELICAL REGION.
 FT DOMAIN 1182 1207 NONHELICAL REGION (C-TERMINAL).
 FT CARBOHYD 56 56 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1354 1354 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT SITE 734 736 CELL ATTACHMENT SITE (POTENTIAL).
 FT SITE 1082 1084 CELL ATTACHMENT SITE (POTENTIAL).
 FT CONFLICT 1450 1450 A -> V (IN REF. 5).
 FT SEQUENCE 1453 AA: 137944 MW: 38802535DF81808 CRC64;

Alignment scores:
 Quality: 450.00 Length: 1143
 Ratio: 0.947 Gaps: 62
 Percent similarity: 41.557 Percent identity: 26.072

Alignment block:

US-09-673-254-1/rev x CALL_MOUSE ..

Align seg 1/1 to: CALL_MOUSE from: 1 to: 1453

2780 CCAGT.....CATACCTCCGAGCTTCTGTGCACGCG 2743
 ||||| ||||| |||||
 167 ProGlyProMetGlyProSerGlyProArgGlyLeuProGlyProProcl 183
 2742 TGCTGGGGA.....GCGGCTCGCGGGCGGCGCTGATC 2708
 ||||| |||||
 183 yAlaProGlyProGlnGlyPheGlnGlyProProGlyGlyLeuProGlyGly 200
 2707 CT.....CACCTCGTCGCGAGCGGTACCCAGGCGCG..... 2676
 ||||| |||||
 200 roGlyGlySerGlyProMetGlyProArgGlyProProGlyProProGly 216
 2675GATGACCCGCGAGCACTCAACGCGCG 2650
 ||||| |||||
 217 LysAsnGlyAspAspGlyGluAlaGlyLysProGlyArgProGlyGlu 233
 2649 A.....CCGTACCGCTACAGCGCGG.....CCAGCGCTACGCGCAGC 2612
 ||||| |||||
 233 gGlyProProGlyProGlnGlyAlaArgGlyLeuProGlyThrAlaGlyL 250
 611 TCCAACAGGCGCACATCATGACCGCGG.....GGAGGCGCG 2574
 ::|||::: |||||
 250 eu...ProGlyMetLysGlyHisArgGlyPheSerGlyLeuAspGlyAla 265
 2573 CAGGCGCTGGCGGACGTCTGGC..... 2550
 ::|||::: |||||
 266 LysGlyAspAlaGlyProAlaGlyProLysGlyGluProGlySerProcl 282
 2549GGTCACTATCACCC..... 2535
 ||||| |||||
 282 yGluAsnGlyAlaProGlyGlnMetGlyProArgGlyLeuProGlyGlu 299
 2534CGGTAGTCCGACCGCGCATCGAGGGGACAGGTG..... 2497
 ||||| |||||
 299 rgGlyArgProGlyPro...ProGlyThrAlaGlyAlaArgGlyAsnAspG 315
 2496CCTGCTCTACTCTCGGCTCAACCCCT 2470
 ||||| |||||
 315 yAlaValAlaGlyAlaAlaGlyProProGlyProThrGlyProThrGlyPro 331
 2469 TCCTGCGTCCGCGGCGAAGGCGCGCACCCCTGTGTGGCTGCGCGTCC 2420

332 .ProGlyPhe..... |||||
 2419 GGGCGGCGCGAGAGGTGACACAGGCGGCTACTACAGGACCGCGCGCT 2370
 ::|||::: |||||
 341 yGlyGlyAlaGlyProGlnGlyAlaArgGlySerGlyLeuProGlnGly 357
 2369 GTCCCGGTGACCGCGCGCGACCGCGCGCGCTCGCGGGAAGCTCT 2320
 ||||| |||||
 358 ValArgGlyGlu...ProGlyProProGlyProAlaGly..... 369
 2319 GGGAGCGCGCGCGCGCGCGCGCGACACCGCGCACTGACGCG..... 2275
 ||||| |||||
 370AlaAlaGlyProAlaGlyAsnProGlyAlaAspGlyGlu 383
 2274 GGGCGCGCGCGCGCGCGCGCGCATGTCTCATCCGCGAGATGCTGCT 2226
 ||||| |||||
 383 oGlyAlaLysGlyAlaAsnGlyAlaProGlyYLeAlaGlyAlaProGly 399
 2225 CATCCGCGAGCGCAGACGCTCTGTCCGATCCATCAAGGAAGCATTC 2176
 ||||| |||||
 400 PheProGlyAlaArg..... 404
 2175 GTACACAGGTTCGCGCGCGCGCGCGCGCATGTTCGACCTGTCGCC 2126
 ||||| |||||
 405GlyProSerGlyProGlnGly.....P 412
 2125 CGATGCGCGCGCGCGCGCG.....ACTTCAACCGCGCTCTTCG 2085
 ||||| |||||
 412 roSerGlyProProGlyProLysGlyAsnSerGlyGluProGlyAlaPro 428
 2084 GCTGAGCGCGCACCGCTGTGTACAGCACCGCGCGCGGAGATACAGCAGC 2035
 ::|||::: |||||
 429 GlyAsnLysGlyAspThrGlyAlaLysGlyGluProGlyAlaThrGlyAla 445
 2034 TTCACCTCCGACGGA..... 2019
 ::|||::: |||||
 445 LcInGlyProProGlyProAlaGlyGluGlnGlyLysArgGlyAlaArgG 462
 2018AGCTGTGCGCGCGGTGCGCGCGCGCGCACAGATCAGCA 1983
 462 LysLysProGlyProSerGlyLeuProGlyProProGlyGlyAlaArgGly 478
 1982 CCCCTTACC.....ACCGCGCTACGCGG 1960
 ||||| |||||
 479 ProGlySerArgGlyPheProGlyAlaAspGlyValAlaGlyProLysG 495
 1959 CCCGGGAACGACACAGCAGCGCATGCCGCGCATCTGAGCGCTACTTGC 1910
 ||||| |||||
 495 yProSerGlyLysArgGlyAlaProGlyProAlaGlyProLysGlySerP 512
 1909 CAGCGACGCGCGCGACGACACTGACCAAGCGGTGCGAGACG..... 1868
 ||||| |||||
 512 ro.GlyGluAlaGlyArgProGlyGluAlaGlyLeuProGlyAlaLysG 528
 1867GGGCGCG 1861
 528 yLeuThrGlySerProGlySerProGlyProAspGlyLysThrGlyProP 545
 1860 GAGGTCAATCATCTCCGATGAGCTCTCGGCTCGG..... 1823
 ||||| |||||
 545 roGlyProAlaGlyGlnAspGlyArgProGlyProAlaGlyProProGly 561
 1822CCGATGGCGGCT 1809
 562 AlaArgGlyGlnAlaGlyAlaMetGlyPheProGlyProLysGlyThrAl 578
 1808 TCGCCGACCCCGCGCGCGCGCGGTTCGCGGTCTG.....GCGCAAGGGA 1765
 ::|||::: |||||
 578 aGlyGlyProGlyLysAlaGlyGlyLysArgGlyLeuProGlyProProGlyAla 595
 1764 GTCATGAGAGGCGCGGAGGTACGCGGCTGCCGCGCTC..... 1727
 ::|||::: ||||| |||||


```

197 ACTGCGCAGCAGGTTCCCGAGCCCGAC.....TGGCCGTCG 160
      ::|||
1135 .....SerIaGlySerProGlyLysAspGlyLeuSnGlyLeuPro... 1148
159 CGTACGACGAGTTGCGGTGTGCGCGA...AGGGGCCAGACGCGCGG 113
1149 .....GlyProIleGlyProProGlyProArgIleYrGrth 1160
112 CTCACGACACTGCGCGCTGGCGCGCTAGTCGCGCGCGCGACGCGG 64
1160 rGlyAspSerGlyProIaGlyProProGlyProProGlyProProGly 1177
      ::|||
63 ..ACGGGACCGCCCGCACC 49
1177 rGProGlyProProSer 1182

seq_name: SwissProt_40:CA13_MOUSE
seq_documentation_block:
CA13_MOUSE STANDARD; PRT; 1464 AA.
P08121; 061429; O9CRN7;
01-AUG-1988 (Rel. 08, Created)
15-JUL-1999 (Rel. 38, Last sequence update)
16-OCT-2001 (Rel. 40, Last annotation update)
Collagen alpha 1(III) chain precursor.
DE COL3A1.
GN Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
      |||
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL X DBA; TISSUE=Embryo;
RX MEDLINE=95011609; PubMed=7926795;
RA Toman D., de Crombrughe B.;
RT "The mouse type-III procollagen-encoding gene: genomic cloning and
RL complete DNA sequence.";
RN Gene 147:161-168(1994).
      |||
RP SEQUENCE OF 1-488 FROM N.A.
RX MEDLINE=88167858; PubMed=3443309;
RA Wood L., Theriault N., Vogeli G.;
RT "Complete nucleotide sequence of the N-terminal domains of the murine
RL alpha-1 type-III collagen chain.";
RN Gene 61:225-230(1987).
      |||
RP SEQUENCE OF 1-28 FROM N.A.
RX MEDLINE=85131189; PubMed=3972847;
RA Liu G., Mudry M., de Crombrughe B.;
RT "Identification of the promoter and first exon of the mouse alpha 1
RT (III) collagen gene.";
RN J. Biol. Chem. 260:3773-3777(1985).
      |||
[4]
SEQUENCE OF 810-1464 FROM N.A.
STRAIN=C57BL/6J; TISSUE=Embryonic head;
MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Aikawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Iwawa M., Nishi K., Kiyosawa H., Kondo S., Yamanka I.,
RA Saito T., Okazaki Y., Gojouri T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochava H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Peesle G., Quackenbush J.,
RA Schirral L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barish G.,
RA Blake J., Boilelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gwincich S., Hill D., Hofmann M., Hume D.A., Kamliya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Rodriguez I., Sakamoto N.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata K., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weltz C., Whitlaker C., Wilmink L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohlsuki S.,

```

```

RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RN [5]
RP SEQUENCE OF 1442-1464 FROM N.A.
RC STRAIN=C57BL;
RX MEDLINE=91274355; PubMed=2054384;
RA Metsaeranta M., Toman D., de Crombrughe B., Vuorio E.;
RT "Specific hybridization probes for mouse type I, II, III and IX
RT collagen mRNAs.";
RL Biochim. Biophys. Acta 1089:241-243(1991).
CC -1- FUNCTION: COLLAGEN TYPE III OCCURS IN MOST SOFT CONNECTIVE TISSUES
CC ALONG WITH TYPE I COLLAGEN.
CC -1- SUBUNIT: TRIMERS OF IDENTICAL ALPHA 1(III) CHAINS. THE CHAINS ARE
CC LINKED TO EACH OTHER BY INTERCHAIN DISULFIDE BONDS. TRIMERS ARE
CC ALSO CROSS-LINKED VIA HYDROXYLYSINES.
CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC O-LINKED GLYCAN CONSISTS OF GLC-GAL DISACCHARIDE (BY SIMILARITY).
CC -1- SIMILARITY: CONSTRAINTS 1 VWFC DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; X52046; CA36279.1; -.
DR EMBL; M18933; AAA3738.1; -.
DR EMBL; K03037; -. NOT ANNOTATED CDS.
DR EMBL; AK019448; BAB31724.1; -.
DR EMBL; X57983; CAA41048.1; -.
DR PIR; A22287; A22287.
DR PIR; A27353; A27353.
DR PIR; S16373; S16373.
DR MCD; MGI:88453; Col3a1.
DR InterPro: IPR000087; Fib.collagen_C.
DR InterPro: IPR001007; VWFC.
DR Pfam; PF01410; COLF1; 1.
DR Pfam; PF01391; Collagen; 17.
DR ProDom; PD002078; Fib.collagen_C; 1.
DR SMART; SM00038; COLF1; 1.
DR SMART; SM00214; VWFC; 1.
DR PROSITE; PS01208; VWFC; 1.
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Glycoprotein; Collagen; Signal.
FT SIGNAL 1 23
FT PROPEP 24 154
FT CHAIN 155 1203
FT PROPEP 1204 1464
FT DOMAIN 31 90
FT DOMAIN 155 169
FT DOMAIN 170 1195
FT DOMAIN 1196 1464
FT CARBOHD 262 262
FT MOD_RES 262 262
FT MOD_RES 283 283
FT MOD_RES 859 859
FT MOD_RES 976 976
FT MOD_RES 1093 1093
FT MOD_RES 1105 1105
FT MOD_RES 1195 1195
FT DISULFD 1196 1196
FT DISULFD 1196 1196
SQ SEQUENCE 1464 AA; 138944 MW; 2104EC27A886090B CRC64;

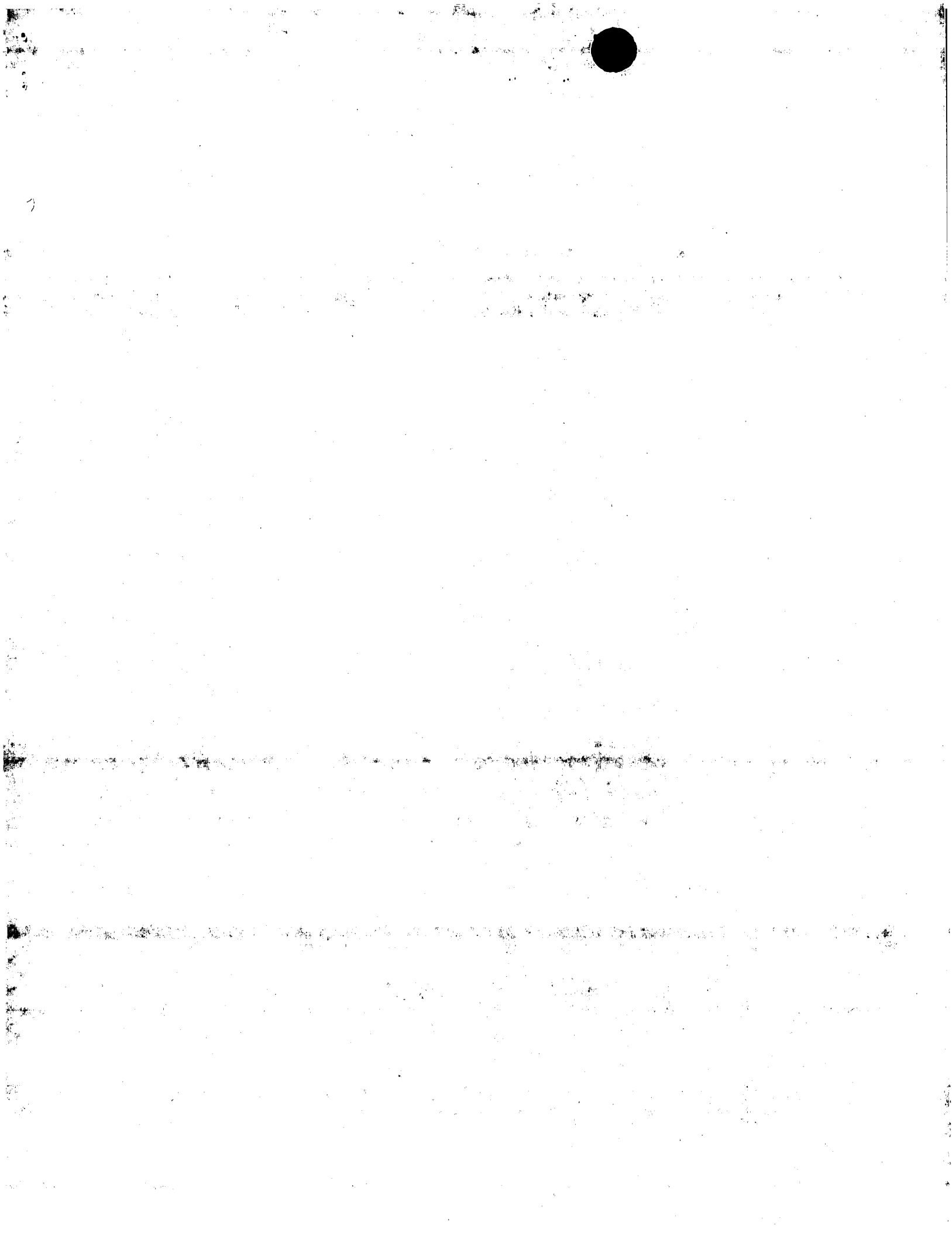
alignment_scores:
Quality: 450.00 Length: 1187
Ratio: 0.911 Gaps: 77
Percent Similarity: 41.618 Percent Identity: 27.043

```



```
1401 TCGGTGTCACTGCGGGAGGGGTACCCCGCGGGCGGGCGGTGCTC 1352
      :          |||||          |||          :|||
681 u.....ArgGlyProProGlyThrAlaGlyIleProG 692
1351 GTGAGGCGGAGGCGGC.....CCGGGTGGCCCTCCAGCC 1317
      ||| |||||
692 LysAlaArgGlyGlyAlaGlyProProGlyProGlyLysGlyPro 708
1316 GTTCGGGTGCCATGATGACATGACGAGCGCAAGCCGAGTGCACGAGC 1267
      :|||          :|||          :|||
709 AlaGlyProProGlyProProGlyAlaSerGlySerProGlyLeuGlnG 725
1266 CCTTCGGGGA.....GGCGGGCCCGGTGTCGAGTGAAGCCCGCGC 1223
      :|||          :|||          :|||
725 yMetProGlyGlyLysArgGlyProGlySerProGly.....ProLysG 740
1222 GGGGAGCCGCGCTGGGTCAATCACCAGTGAACGCCCTCGCCGAGGTGCT 1173
      :|||          :|||          :|||
740 LysIle.....LysGlyIleProGlyGlyAla 748
1172 GGGCGATCC.....CCGGTTCGTGAAGACCCGA..... 1143
      ||| :|||          :|||          :|||
749 GlyAlaAspGlyAlaProGlyLysAspGlyProArgGlyProAlaGlyPr 765
1142 .....CCTCGCCCGCGCGCTGGCGGG.....GTG 1115
      ||| ||| |||
765 oIleGlyProProGlyProAlaGlyIleProGlyAspLysGlyLysIle 782
1114 GACGAGGGCT.....CGACATCCCGTCCGAGAGTGGCTC 1077
      ||| :|||          :|||          :|||
782 LysSerProGlyLeuProGlyIleAlaGlyProArgGlyLysGlyProGly 797
1076 GTTCACGCTCATCGCCCGTGAAGCGGAGCGCCAGCGGCGCTGCGCGCA 1027
      :|||          :|||          :|||
798 .....GluArgGlyGlyIleHisGlyProProGlyProAlaGlyPh 810
1026 TCCAGCCACTGCGTCA..... 1008
      ||| :|||
810 eProGlyAlaProGlyGlnAsnGlyIleProGlyAlaLysGlyLysArg 827
1007 .....CCGCGCGCGGTGGCGCA 990
      :|||          :|||          :|||
827 LysAlaProGlyGlyLysGlyGlyGlyLysProProGlyProAlaGlyPro 843
989 GCGGAGGATCGCATCGCGCGATCGCGCGGCTGCTACCGCACTCG 940
      :|||          :|||          :|||
844 ThrGlySerSer.....GlyProAlaGly..... 851
939 CCGAGCCCTCCGCGCGGTGGGCAAAACCGCGAGCTGATCGCGGCTTC 890
      ||| ||| |||
852 .....ProProGlyProGlnGlyAlaLysGlyLysArg..... 862
889 GCGTACCACTCCGCTGTGGTCACTGAGCTGCTGCTGGTGGCGGT 840
      :|||          :|||          :|||
863 .....GlySerProGlyGlyProGlyThrAlaGlyPheProGly 875
839 CACCGATCCGGGATGCGCCGAGAGCGGTCAAGCTTCAAGCACTCG 790
      :|||          :|||          :|||
876 GluArgGlyLeuProGlyProProGlyAsnAsnGlyAsnProGly..... 890
769 GCGTCGGGCGCGGAGAGCGGGGAGCGAGCGACCGCTGCGCGG 740
      :|||          :|||          :|||
891 ..ProProGlyProSerGlyAlaProGly.....LysAspGlyPro...ProG 905
739 GCGGTGCGGACACCTCGC.....CCTGAGAGCT.....GCTCT 702
      :|||          :|||          :|||
905 LysProAlaGlyLysSerGlySerProGlyAsnProGlyIleAlaGlyPro 921
701 CGAACCCTGCACTACGCCGCGGAGACGACACCCGA..... 664
      :|||          :|||          :|||
922 LysGlyAspAlaGlyIleProGlyLysGlyProProGlyAlaGlnG 938
663 .....CATGACCC.....GCGTGTCTACGAGCGCGC 636
```

```
938 yProProGlySerProGlyProLeuGlyIleAlaGlyLeuThrGlyAla 955
      :|||          :|||          :|||
635 GCAGCGCGAGTGGCTCGGTCTCCGACGACGAGCTGCTACATGATCA 586
      :|||          :|||          :|||
955 rg.....GlyLeuAlaGlyProProGlyLysMetProGlyProArgGlySer 969
585 CCGGCG.....TCATCTTCGCGCGGACGACA... 559
      :|||          :|||          :|||
970 ProGlyProGlnGlyIleLysGlyLysSerGlyLysProGlyAlaSerG 986
558 .....CCACCGGCTCTTCCTGAGCTTCCTGCTCGC 528
      :|||          :|||          :|||
986 LysAsnGlyLysArgGlyProProGlyProGln..GlyLeuProGlyGln 1002
527 GAGGTCTTCGGCGCGCGCT...CGCGCGGATGCCGACGAGAGCGCGC 481
      :|||          :|||          :|||
1003 ProGlyThrAlaGlyIleProGlyLysAspGlyAsnProGlySerAspG 1019
480 TCTCCGGTTCGTGGAGAGGCGCTGCGCTACCAACCGCGCGTGGCTAC 431
      :|||          :|||          :|||
1019 yGlnProGlyLysAspGlySer.....ProGlyGlyLys.... 1030
430 ACGTTGTGAGGTTCGCTGCCACGAGGTGACCATCGCGCGCTCGGCT 381
      :|||          :|||          :|||
1031 .....GlyAsp...ArgGlyLysAsnGly 1037
380 GCGCCCGGAGCGCGCGTCTGTGTCAT..... 351
      :|||          :|||          :|||
1038 SerProGlyAlaProGlyAlaProGlyHisProGlyProProGlyProVa 1054
350 .....CGAGGCGACCAACCGAGCGCGC 326
1054 IeGlyProSerGlyLysSerGlyAspArgGlyIleThrGlyProAlaGlyP 1071
325 CATCACGAGCGCCC.....GCAAGCTTCACCGCGAGCTC. 289
      :|||          :|||          :|||
1071 roSerGlyAlaProGlyProAlaGlyAlaArgGlyAlaProGlyProGln 1087
288 .....CCTGTGGGCGC..... 277
      :|||          :|||          :|||
1088 GlyProArgGlyAspLysGlyIleThrGlyLysArgGlySerAsnGlyIle 1104
276 .....GGCTCACTT 267
      :|||          :|||          :|||
1104 eLysGlyHisArgGlyPheProGlyLysnProGlyProProGlySerProG 1121
266 CCGCGAGCGCGCGCACTACTGAGGAGACAGCACTGCCAGCTGAGCT 217
      :|||          :|||          :|||
1121 LysAlaAlaGlyHisGlnGlyAlaIleGlySerProGlyProAlaGlyPro 1137
216 CCGCAGCATGATCGGCGCTAC.....TGGAGGACAG 185
      :|||          :|||          :|||
1138 ArgGlyProValGlyIleProHisGlyProProGlyLysAspGlyThrSerG 1154
184 TTCGCCGAGCGCGACGTGCGCGGTACGACGAGTTGCGTGGTGGCG 135
      :|||          :|||          :|||
1154 yHisProGlyProIleGlyPro..... 1161
134 GAAGGCGCGCCAGA.....CGCGCGGCTCACCGAGCTGCCG 97
      :|||          :|||          :|||
1162 ..ProGlyProArgGlyLysnArgGlyLysArgGlyLysSerIleLysSerPro 1177
96 TCT.....GGCTGCGCTGATCCGCGCGGCGGACCGGACGAGCGCGC 53
      :|||          :|||          :|||
1178 GlyHisProGlyGlnProGlyProPro.....GlyProPr 1189
52 CACCGCGACCGCGCGGTGGCGGTCCGTCGCTGATACCGGTCGCGCA 4
      :|||          :|||          :|||
1189 o.....GlyAlaProGlyProCys.....CysGly 1197
```

OM of: US-09-673-254-1 to: SPTREMBL_19:* out_format : pfs

Date: Jun 11, 2002 8:19 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

```
-MODEL=framed_n2p.model -DEV=xlh  
-O=/cgml2/USPTO.spool/US09673254/runcat_11062002.114213.1012/app_query.fasta_1.2973  
-DB=SPTREMBL_19 -OFMT=fasta -SUFFIX=n2p -SPT -GAPOP=12.000  
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPEL=0.000 -LOOPEXT=0.000  
-OGAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500  
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500  
-OGOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blonsum62  
-US-human40.cdi -LIST=45 -DOCCALIGN=200 -THR_SCORE=Pct  
-MAX=100 -THR_MIN=0 -ALIGN=7 -MODE=LOCAL -OUTFMT=pfs  
-M-ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USFR=US09673254.ecgml_1360 -NCPU=6 -ICPU=3 -LONGLOG  
-DEV=TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPEX -WAIT -THREADS=1
```

Search information block:

```
Query: US-09-673-254-1  
Query length: 2870  
Database: SPTREMBL_19:*  
Database sequences: 562222  
Database length: 172994929  
Search time (sec): 134.250000
```

score_list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
SP_bacteria:Q9ZAU3	- 2171.00	1722.23	3.3e-88	415	09ZAU3 streptomyces peucetius.
SP_bacteria:Q59971	- 2087.00	1655.77	1.6e-84	422	059971 streptomyces sp. cytoch
SP_bacteria:Q93M12	- 2062.00	1636.07	2.0e-83	420	093M12 streptomyces peucetius
SP_bacteria:Q9ZAU2	- 1465.00	1168.27	3.6e-57	275	09ZAU2 streptomyces peucetius
SP_bacteria:Q50578	- 1385.00	1105.11	1.2e-53	275	050578 streptomyces sp. orfa p
SP_bacteria:Q9ZAU1	- 1014.00	811.84	2.4e-37	287	09ZAU1 streptomyces peucetius
SP_bacteria:Q53877	- 980.00	785.00	7.6e-36	287	053877 streptomyces sp. c5. ket
SP_bacteria:Q55077	- 975.00	784.27	1.2e-35	195	055077 streptomyces sp. daunom
SP_virus:Q99307	- 538.00	426.28	2.3e-16	924	099307 epstein-barr virus (stra
SP_vertebrate:Q9Y1B4	- 506.00	397.25	6.0e-15	1450	09Y1B4 cynops pyrrhogaster (ja
SP_virus:Q905K9	- 486.50	389.11	4.0e-14	608	0905K9 herpesvirus papio. ntr.
SP_human:Q60382	- 486.50	380.09	4.4e-14	1791	060382 homo sapiens (human). x
SP_human:Q60382	- 484.50	374.93	5.5e-14	2752	090382 homo sapiens (human). x
SP_human:Q15038	- 484.00	378.16	5.6e-14	1783	015038 homo sapiens (human). x
SP_human:Q15038	- 479.00	374.21	9.3e-14	1783	015038 homo sapiens (human). x
SP_human:Q90Q35	- 479.00	374.17	9.3e-14	1791	060382 homo sapiens (human). x
SP_human:Q90Q35	- 479.00	370.59	1.0e-13	2752	090382 homo sapiens (human). x
SP_human:Q90H48	- 478.00	371.31	1.0e-13	2296	090H48 homo sapiens (human). s
SP_vertebrate:Q93251	- 476.50	373.99	1.2e-13	1445	093251 rana catesbeiana (bull)
SP_invertebrate:Q9VPG1	- 475.50	369.38	1.4e-13	2284	09VPG1 drosophila melanogast
SP_human:Q90Q36	- 473.00	372.27	1.7e-13	1275	090Q36 homo sapiens (human). x
SP_bacteria:Q9K9K7	- 472.50	385.05	1.6e-13	263	09K9K7 streptomyces coelicolor
SP_virus:Q41971	- 470.00	374.07	2.2e-13	774	041971 murid herpesvirus 4. hyf
SP_human:Q90Q39	- 468.00	370.73	2.7e-13	956	090Q39 homo sapiens (human). x
SP_bacteria:Q9F200	- 467.00	376.55	2.8e-13	433	09F200 streptomyces coelicolor
SP_virus:Q41971	- 460.50	366.57	5.7e-13	774	041971 murid herpesvirus 4. hyf
SP_human:Q90Q40	- 459.00	361.30	6.9e-13	1262	090Q40 homo sapiens (human). x
SP_vertebrate:Q91B91	- 458.50	359.77	7.3e-13	1447	091B91 xenopus laevis (afrikan
SP_human:Q76045	- 458.00	359.29	7.7e-13	1461	076045 homo sapiens (human). f
SP_rodent:Q63079	- 454.50	356.58	1.1e-12	1453	063079 rattus norvegicus (rat)
SP_invertebrate:Q91365	- 454.00	352.60	1.2e-12	2232	091365 caenorhabditis elegat
SP_human:Q90H48	- 452.50	351.18	1.4e-12	2296	090H48 homo sapiens (human). s
SP_human:Q90Q39	- 452.00	356.09	1.4e-12	956	090Q39 homo sapiens (human). x
SP_virus:Q41972	- 451.50	360.23	1.4e-12	706	041972 murid herpesvirus 4. hyf
SP_virus:Q41973	- 451.00	359.59	1.5e-12	727	041973 murid herpesvirus 4. hyf
SP_human:Q90Q40	- 450.50	354.59	1.6e-12	1262	090Q40 homo sapiens (human). x
SP_human:Q90Q36	- 449.00	353.32	1.9e-12	1275	090Q36 homo sapiens (human). x
SP_virus:Q99307	- 448.50	357.62	2.0e-12	924	099307 epstein-barr virus (stre
SP_bacteria:Q52569	- 444.50	357.40	2.8e-12	511	052569 amycolatopsis mediterr
SP_virus:Q905K9	- 443.50	353.16	3.1e-12	608	0905K9 herpesvirus papio. ntr.

```
SP_virus:Q41972 + 437.50 349.18 5.8e-12 706 041972 murid herpesvirus 4.  
SP_vertebrate:Q90YJ0 - 436.00 342.57 7.1e-12 1352 090YJ0 brachydanio rerio (z  
SP_virus:Q69340 - 434.50 338.30 8.5e-12 1958 069340 pseudorabies virus.  
SP_human:Q9UFE83 - 430.50 345.29 1.2e-11 580 09UFE83 homo sapiens (human).  
SP_invertebrate:Q9VE45 - 430.00 340.78 1.3e-11 950 09VE45 drosophila melanoga
```

seq_name: SP_bacteria:Q9ZAU3

seq_documentation_block:

```
ID 09ZAU3 PRELIMINARY: PRT: 415 AA.  
AC 09ZAU3:  
DT 01-MAY-1999 (TREMBLrel. 10. Created)  
DT 01-MAY-1999 (TREMBLrel. 10. Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17. Last annotation update)  
DE CYTOCHROME P450 129A2 (DAUNORUBICIN C-14 HYDROXYLASE).  
GN CYP129A2 OR DOXA.  
OS Streptomyces peucetius.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.  
OX NCBI_TaxID=1950;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 29050;  
RX MEDLINE=99084971; PubMed=9864344;  
RA Lomonoskaya N., Otten S.L., Doi-Katayama Y., Fongstein L., Liu X.-C.,  
RA Takatsu T., Invenit A., Filippi S., Torti F., Colombo A.L.,  
RA Hutchinson C.R.;  
RT "Doxorubicin overproduction in Streptomyces peucetius: cloning and  
RT characterization of the dnru ketoreductase and dnrv genes and the doxa  
RT cytochrome P-450 hydroxylase gene."  
RL J. Bacteriol. 181:305-318(1999).  
CC -1 SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.  
DR EMBL: U77891; AAD04715.1; -  
DR HSSP: 000441; 10XA.  
DR InterPro: IPR001128; Cyt_P450.  
DR Pfam: PF00067; P450. 2.  
DR PRINTS: PR00385; P450.  
DR PROSITE: PS00086; CYTOCHROME_P450; UNKNOWN_1.  
KW Heme; Monooxygenase; Oxidoreductase.  
SQ SEQUENCE 415 AA; 45095 MW; CCF873824BEB6CCF CnC64;
```

alignment_scores:

```
Quality: 2171.00 Length: 415  
Ratio: 5.231 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 99.759
```

alignment_block:

US-09-673-254-1/rev x Q9ZAU3 ..

Align seg 1/1 to: Q9ZAU3 from: 1 to: 415

```
1330 GTGGCGGTGACCGCGTTGCGGTGCGCATGATGACGAGCGCAAGC 1281  
:::|||||  
1 MetAlaValAspProPheAlaCysProMetMetThMetInArgLysP 17  
1280 CGAGGTGACGACGCGCTCCGGAGGCGCGCGCGTGGAGGTGAAG 1231  
|||||  
17 OGtUvAlHISAspAlaPheArgGluAlaGlyProValGluValAsn 34  
1230 CCCCCGCGGCGGACCGCGCTGGGTATCATCAGCATGACGCGCGCCG 1181  
|||  
34 LArProAlaGlyGlyProAlaTrpValIleThrAspAlaLeuAlaArg 50  
1180 GAGGTGTCGCGCATGCCGCGTTGCGGAAGAGCGGACCTGCGCCCGC 1131  
|||||  
51 GluValLeuAlaAspProArgPheValIlyAspProAspLeuAlaProAl 67  
1130 CGCGTCGCGGCGGTGAGAGCGGCTCGACATCCCGTTCCGAGACTGC 1081  
|||||  
67 AlaTrpArgGlyValAlaAspArgLysLeuAspIleProValProGluLeu 84  
1080 GTCCGTTCAACGCTCATCGCGGTGAGCGGCGGACCGCCACGCGCGC 1031
```

```

|||||
84 rgProPheThrLeuLeuAlaValAspGlyGluAlaHisArgArgLeuArg 100
1030 CGCATCCACGACCTGGCTTCAACCGCGCGCGCGCGAGACGAGCA 981
101 ArgIleHisAlaProAlaPheAsnProArgArgLeuAlaGluArgThrAs 117
980 TCGCATCGCGCGCATCGCGCGCGCTGCTCAACGCACTCGCGAGCGCT 931
117 PArgIleAlaAlaIleAlaGlyArgLeuLeuThrGluLeuAlaAspAla 134
930 CGCGCGCGCTGGGCAACCGCGCGAGCTGATCGCGCGCTTCGCGTAC 881
134 erGlyArgSerGlyLysProAlaGluLeuIleGlyGlyPheAlaTrpHis 150
880 TTCGCGCTGGCTGATCGCGAGCTGCTCGCGTGGCGCGTACCGCATCC 831
151 PheProLeuLeuValIleCysGluLeuLeuGlyValProValThrAsp 167
830 GCGCATGGCGCGCGAGCGCGTTCAGCGCTTCAAGCACTCGCGCTCG 781
167 CalAmetAlaArgGluAlaValSerValLeuLysAlaLeuGlyLeuGly 184
780 GCCCGCAGACGCGCGGGGTGACGCGACGACCTGCGCGGGCGCTGCC 731
184 LyrProGlnSerGlyGlyAspGlyThrAspProAlaGlyValPro 200
730 GACACCTCGCGCGTGGAGACCTGCTCGCTCGAAGCGCGTCACTCG 681
201 AspThrSerAlaLeuGlnSerLeuLeuLeuGluAlaValHisSerAla 217
680 GCGGAGACGACCCCGACCATGACCGCGCTGCTGTAAGACGCGCGCAG 631
217 gArgAsnAspThrProThrMetThrArgValLeuTrpGluArgAlaGln 234
630 CGGATTCGGCTCGCTTCCGACGACGACCTGCTACTGATGACCGCG 581
234 IagLpHeGlySerValSerAspAspGlnLeuValTrpMetIleThrGly 250
580 CTGATCTTCGCGCGCGCACGACACCGCGCTCTTCTGCGCTCTGCT 531
251 LeuIlePheAlaGlyHisAspThrThrGlySerPheLeuGlyPheLeu 267
530 CGCGGAGCTCTGCGCGCGCGCTCGCGGAGATGCCAGACGACGCG 481
267 uAlaGluValLeuAlaGlyArgLeuAlaAlaAspAlaAspAlaVal 284
480 TCTCCGGTCTGGAGGAGCGCTGCGCTACCGCGCGCGCTGCGCTAC 431
284 alSerArgPheValGluGluAlaLeuArgTrpHisProProValPro 300
430 ACCTTGTGAGAGTCTGCTGCCACGAGGATGACCATCGCGCGCTCGCT 381
801 ThrLeuTrpArgPheAlaAlaThrGluValThrIleGlyGlyValArg 317
380 GCCCGCGGAGCGCGCTGCTGAGACATCGAGGCGACCAACACCGCAG 331
317 uProArgGlyAlaProValLeuValAspIleGluGlyThrAsnThrAsp 334
330 GCGCGCATACGAGCGCGCGCGCTTCCACCGAGCGCTCCCTCGTGG 281
334 LyrArgHisHisAspAlaProHisAlaPheHisProAspArgProSer 350
280 CGCGCGCTACCTTCGCGCGAGCGCGCACTACTGATCGCGGAGACGCT 231
351 ArgArgLeuThrPheGlyAspGlyProHisTrpCysIleGlyGluGln 367
230 CGCCAGCTGAGTGGCGCAGCATGATGCGCTACTGCGCAGCAGAGTTC 181
367 uAlaGlnLeuGlnSerArgThrMetIleGlyValLeuArgSerArgPhe 384
180 CGGAGCGCGCTGCGCGCTGCGAGAGTTCGCGTGGTGGCGCGAG 131
|||||

```

```

384 roGluAlaArgLeuAlaValProTrpAspGluLeuArgTrpCysArgLys 400
130 GGGGCCAGACGCGCGCGCTGACCGAGCTCCGCTGCGCTGCGC 86
401 GlyAlaGlnThrAlaArgLeuThrGluLeuProValThrPleuArg 415
seq_name: sp_Bacteria:Q59971

```

```

seq_documentation_block:
ID Q59971 PRELIMINARY; PRT; 422 AA.
AC Q59971;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE CYTOCHROME P450 129A1 (DAUNOMYCIN C-14 HYDROXYLASE).
GN CYP129A1 OR DOXA.
OS Streptomyces sp.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1931;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CS;
RX MEDLINE=96236066; PubMed=8655529;
RA Dickens M.L., Strohl W.R.;
RT "Cloning, sequencing, and analysis of aklaviketone reductase from
Streptomyces sp. strain C5."
RL J. Bacteriol. 178:3384-3388(1996).
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR EMBL; U50973; AAB08049.1; -.
DR HSSP; Q00441; IOXA.
DR InterPro; IPR011128; Cyt_P450.
DR Pfam; PF00067; P450; 2.
DR PROSITE; PS00086; CYTOCHROME P450; UNKNOWN_1.
KW Electron transport; Heme; Membrane; Monooxygenase; Oxidoreductase.
FT BINDING 369 369; HEME (BY SIMILARITY).
SQ SEQUENCE 422 AA; 46096 MW; F4EDECAD1D39052 CRC64;

```

```

alignment_scores:
Quality: 2087.00 Length: 422
Ratio: 5.041 Gaps: 0
Percent Similarity: 98.104 Percent Identity: 94.550
alignment_block:
US-09-673-254-1/rev x Q59971 ..

```

Align seg 1/1 to: Q59971 from: 1 to: 422

```

1351 GTGAGCGGCGAGCGCGCGGCTGGCGCTGACCGCTTCCGCTGCCAT 1302
1 MetSerGlyGluAlaProArgValAlaValAspProPheSerCysPro 17
1301 GATGACCATGAGCGCGCAACCGCGAGGTGACGAGCGCTTCGGGAGCGG 1252
17 LweThrMetClnArgLysProGluValHisAspAlaPheAlaGluAlaG 34
1251 GCCCGGTCTGAGGTGAACGCCCGCGCGGCGGAGACCGCGCTGGGTCA 1202
34 LyrProValAlaGluValAsnAlaProAlaGlyGlyProAlaTrpValIle 50
1201 ACCGATGACCGCTGCGCGGAGGTGCTGCGCATCCCGCTGCTGTGTA 1152
51 ThrAspAspAlaLeuAlaArgGluValLeuAlaAspProArgPheVal 67
1151 GAGACCGGACCTTCGCGCGCGCGCGCTGCGGCGGCGGTGACGAGCTTC 1102
67 SasProAspLeuAlaProThrAlaTrpArgGlyValAspAspGlyLeu 84
1101 ACATCCCGCTTCGAGAGCTGCTCGTTCACCGCTCATCGCTGAGCGC 1052
84 SplLeProValProGluLeuArgProPheThrLeuIleAlaValAspGly 100

```


1051	GAGGGGCAACGGGGCGCTGGCGGGCATCCAGCCACTGGGCTTCAACCGCG	1002
101	GIUASPHISATGADGULEATGATGATGTTGTTTSAIPROALAPHEASNPROT	117
1001	CCGGCTGGCGCGGCGAGCATTCGATCGCGCGCATCGCGCGCGCGCTGC	952
117	GATGULENLAAGTATGTTTATPARGATTAALALALALALALALAPRATGUL	134
951	TCACCGAATCGCGAGCGCTCCGGCGCGTCCGGGCAACCGCGCGAGCTG	902
134	EUTHTGULUENLAASPERSEATPARGSERGTGULPROALAGTULUEN	150
901	ATCGCGCGGCTTCCGCTACCACTTCCCGCTGTGGTCATCTCGAGAGTCT	852
151	TTGGTGTGULPHALATATYTHISPHETROUENULALIECTSGTULUEN	167
151	CGGTTGCGCGGCGACGATTCGGGATGCGCGGAGCGCGCGCATCGCTTC	802
167	UGULYALAPROVALTPHAPPROALAMEALALATGULALVALALYAL	184
801	TCGAAGGCACTCGGCTCGGCGCGCGCGAGCGCGGCGGTGAGCGGACG	752
184	EULYALALUENULYUENUGLYTPROGINSERTAGLYLASYRGLYTHC	200
751	GACCTGCGCGGGGGCGTCCCGGACACTCGGCGCTGGAGAGCTTGCTCT	702
201	ASPPROALAGLYASPYVALPPOASPHRISERATLEUENUSERTLEULE	217
701	CGAAGCGCGGCACTAGCGCGCGCGGAGACAGACCCCGGCACTGACCGCG	652
217	UGULALVALHISALALALATGULYASPRHTARGTTHMETHTHARGV	234
651	TCGCTGACGAGCGCGCGAGCGCGGATTCGCGCTCGCGAGGACGACG	602
234	ALUENUTGULATGATAGTALAGTALAGTUPHEGTYSERVALISERASPRALN	250
601	CTCGCTACATGATATACCGGGCTATCTTCCCGCGGACGACACACCGG	552
251	LEUVALTYMETLTETHTGTYLEUILLERHEALAGTUNHISAPRHTHTGTL	267
551	CTCCTTCGCGGCTTCGCTGCTCGCGAGCTCTCGCGCGCGCGCTCCGCG	502
267	YSETPHEUENUGLYPHELEUENALAGTUNALYUENALUATGATGULEALAA	284
501	CGGATTCGCGAGGAGGACCGCTTCCCGCTCCGCTTCGTHGAGAGGCGCTCGC	452
284	IASPRALASPRULYASPRALATIESERTARGPHEVALGULALALUENATG	300
451	THCSACCGCGCGCTCCCTACACGTTGTGGAGTTCGCTGCCACGAGAT	402
301	HNASHISPROFOVALPRTOLYTHITLEUTPRATGPRHEALALATATHTGULVA	317
401	GACGATTCGCGCGGCTCGCGCTCCCGCGGAGGCGCGGTGTGGTGACA	352
317	TAALLIETARGLYVALATGULEPRTOLATGULALATROVALLEUVALASPR	334
351	TCGAGGCGCACCAACGAGCGGCGCGCATCTACGAGCGCGCGCGCGCTTC	302
334	LEGLUGLYTHASPHRHSASPRGLATGTHSHIASPRALATROHISALAPHE	350
301	CACCGCGAGCGCTCTGTCGCGCGGCTCACTTTCGGGACGAGGCGGCA	252
351	HISPROASPRATGPROSERATGATGATGATGULHTHRPHEGLYASRGLYPRONH	367
251	CPACGAGCATCGGGGAGAGAGCTCGGCGCATCTGAGAGTGGCGGCACTGATCG	202
367	STATGYSITTEGTYGULINLEUENALAGTUNLEUGLUSERTATGTHMETTIEG	384
201	GGGATCTGGCGAGAGGATTCGCCGAGGCGCGCATCTGCGCGCTTACGAC	152
384	LYVALLEUATGSEATATGPRHERPROGINALATATGULEALALVALATPOTGULU	400
151	GAGTTTGCGGTGCTGCCGAGAGGGGCGCGAGACGCGCGGCTCACCGAAT	102

```

|||||
401 GULENATGTRCYsarglySGlyAlaGlnThrAlaArgLeuThrAspLe 417
|||||
101 GCCCGCTGCGCTGCGC 86
|||||
417 urovatlrpleuArG 422

seq_name: sp_bacteria:Q93MI2

seq_documentation_block:
ID Q93MI2 PRELIMINARY: PRT: 420 AA.
AC Q93MI2;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE DAUNOMYCIN C-14 HYDROXYLASE.
GN DOXA.
OS Streptomyces peucetius subsp. caesius.
OC Bacteria; Firmicutes; Actinobacteriae; Actinobacteriidae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OC NCBI_TaxId=55158;
[1]
RP
SEQUENCE FROM N.A.
RC STRAIN=ATCC 27952;.
RC Hong Y.-S., Kim H.-S., Lee J.-H., Kim K.-W., Lee J.-J.;
RA Molecular Cloning and Characterization of the dox A Cytochrome P-450
RT Hydroxylase Gene in Streptomyces peucetius subsp. caesius ATCC
RT 27952. ";
RL
DR J. Microbiol. Biotechnol. 0:0-0(2001).
EMBL AF403708; AAK95626.1; -.
SO SEQUENCE 420 AA; 4566 MW; 2083C9FE9BA802C6 CRC64;
```

alignment_scores:	Quality: 2062.00	Length: 422
Ratio: 4.993	Gaps: 2	
Percent Similarity: 97.867	Percent Identity: 94.313	
alignment_block:		
US-09-673-254-1/rev x Q93MI2	..	
Align seg 1/1	to: Q93MI2	from: 1 to: 420
1351 GTGAGGGCGAGGCGCCCGGGTGGCCGTGCACCCGTTCCGTTGCCAT	1302	
1352	1303	
1353	1304	
1354	1305	
1355	1306	
1356	1307	
1357	1308	
1358	1309	
1359	1310	
1360	1311	
1361	1312	
1362	1313	
1363	1314	
1364	1315	
1365	1316	
1366	1317	
1367	1318	
1368	1319	
1369	1320	
1370	1321	
1371	1322	
1372	1323	
1373	1324	
1374	1325	
1375	1326	
1376	1327	
1377	1328	
1378	1329	
1379	1330	
1380	1331	
1381	1332	
1382	1333	
1383	1334	
1384	1335	
1385	1336	
1386	1337	
1387	1338	
1388	1339	
1389	1340	
1390	1341	
1391	1342	
1392	1343	
1393	1344	
1394	1345	
1395	1346	
1396	1347	
1397	1348	
1398	1349	
1399	1350	
1400	1351	
1401	1352	
1402	1353	
1403	1354	
1404	1355	
1405	1356	
1406	1357	
1407	1358	
1408	1359	
1409	1360	
1410	1361	
1411	1362	
1412	1363	
1413	1364	
1414	1365	
1415	1366	
1416	1367	
1417	1368	
1418	1369	
1419	1370	
1420	1371	
1421	1372	
1422	1373	
1423	1374	
1424	1375	
1425	1376	
1426	1377	
1427	1378	
1428	1379	
1429	1380	
1430	1381	
1431	1382	
1432	1383	
1433	1384	
1434	1385	
1435	1386	
1436	1387	
1437	1388	
1438	1389	
1439	1390	
1440	1391	
1441	1392	
1442	1393	
1443	1394	
1444	1395	
1445	1396	
1446	1397	
1447	1398	
1448	1399	
1449	1400	
1450	1401	
1451	1402	
1452	1403	
1453	1404	
1454	1405	
1455	1406	
1456	1407	
1457	1408	
1458	1409	
1459	1410	
1460	1411	
1461	1412	
1462	1413	
1463	1414	
1464	1415	
1465	1416	
1466	1417	
1467	1418	
1468	1419	
1469	1420	

```

117 gArlGleuAlaGluArgThrAspArgIleAlaAlaIleAlaGlyArgLeu 134
951 TCACCGAAGTCCGCGACGCTCCGGCGGTCGGGCAACCGCGCGAGCTG 902
134 eutHrCluLeuAlaAspAlaSerGlyArgSerCluGluProAlaGluLeu 150
901 ATCGGCGGCTTCGGCTACCACTTCGCGTGTGGTCATCTGCAGAGTGT 852
151 lIeGlyGlyPheAlaThrHisPheProLeuValIleGlySerIleLeu 167
851 CGGTGGCGGTCACCGATCCGCGGATGGCCGCGGAGCGCGCTCAGGCTTC 802
167 uGlyValProValThrValPro...MetAlaArgGluAlaValSerVal 183
801 TCAGGCACTCGGCTCGCGCGCGCGCGAGAGCGCGGCGGTGACGCGCAG 752
183 euLysAlaLeuAla...SerAlaAlaGlnSerCluGlyGlyAspGlyThr 198
751 GACCTCGCGCGGCGGCGCGCGAGACCTCGCGCTGGAGAGCCGTGCTCT 702
199 AspProAlaGlyGlyValProAspThrSerAlaLeuGlnSerLeuLeu 215
701 CGAAGCGCTGACACTGACCGCGCGGAGACACACCCGACCATGACCGGG 652
215 uGluAlaValHisSerAlaArgArgAsnAspThrProIleMetThrArgV 232
651 TCCTGTACAGCGCGCGGAGCGGAGTTCGCGTCTCCGACGACGACG 602
232 alLeuTyrCluHisThrGlnAlaGluPheGlySerValSerAspAsnGln 248
601 CTCGCTCATATGATACCGCGGCTCATCTTCGCGCGGCGACGACACCGG 552
249 LeuValTyrMetIleThrGlyIleIlePheAlaGlyHisGluArgThrCl 265
551 CTCCTTCCTCGGCTTCCTGCTCGCGAGGTCCTGGCGGCGCGCTCGGCG 502
265 ySerPheLeuGlyPheLeuLeuAlaGlyValLeuAlaGlyArgLeuAla 282
501 CGGATGCCGACGAGGAGCGCGCTCTCCCGGTCGTGAGAGGCGCTGCCG 452
282 lAspAlaAspIleuAspAlaValSerArgPheValGluGluAlaValArg 298
451 TACCAACCGCGCGGCTGCTACACGTTGTGAGGTTGCTGCCACGAGGT 402
299 TyrHisProProValProTyrThrLeuTyrArgPheAlaAlaThrCluVa 315
401 GACCATCGGCGGCTCGGCTGCGCGCGGAGCGCGGCTGGTGAGCA 352
315 lThrIleGlyGlyValArgLeuProProGlyAlaProValLeuValAspI 332
351 TCGAGGGGACCAACACCGAGCGCGCCCATCACAGCGCGCGCGCGCTTC 302
32 lGlyGlyThrAsnThrAspGlyArgGlnHisAspAlaProHisAlaPhe 348
01 CACCGGAGCCGCTCCGCGGCGGCGGCTCACCTTCGCGGAGCGCGCGA 252
349 HisProAspArgProSerTyrArgArgLeuThrPheGlyAspGlyProH 365
251 CTACTGATCGGCGGAGAGCTCGCGCAGCTGAGTGGCGCAGAGATGATG 202
365 sTyrCysIleGlyGlyGlnLeuAlaGlnLeuGlnSerArgThrMetIleG 382
201 GCGTACTCGGACAGGTTCCCGAGGCGCGCATGCGCGTGCCTACGAC 152
382 lValLeuArgSerArgPheProGluAlaArgLeuAlaValProTyrAsp 398
151 GAGTGGCGGTGTGCGCGAGGCGCGGCGGCGCGCTACCGGACT 102
399 GluLeuArgTyrCysArgAsnGlyAlaGlnThrAlaArgLeuThrGluLe 415
101 GCGCGTGTGCGCGC 86
415 uProValIleuArg 420

```

```

seq_name: sp_bacteria:Q9ZAU2
seq_documentation_block:
ID Q9ZAU2 PRELIMINARY; PRT; 275 AA.
AC Q9ZAU2;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE DOXORUBICIN BIOSYNTHESIS ENZYME DNRV.
GN DNRV.
OS Streptomyces peuceletius.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1950;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 29050;
RX MEDLINE=99084971; PubMed=9864344;
RA Lomovskaya N., Otten S.L., Doi-Katayama Y., Feinstein L., Liu X.-C.,
RA Takatsu T., Invenit A., Filippi S., Torti F., Colombo A.L.,
RA Hutchinson C.R.;
RT "Doxorubicin overproduction in Streptomyces peuceletius: cloning and
RT characterization of the dnrV ketoreductase and dnrV genes and the doxa
RT cytochrome P-450 hydroxylase gene."
RL J. Bacteriol. 181:305-318(1999).
DR EMBL; U77891; AB004716.1; -.
DR InterPro: IPR004360; Gly_bleo_diox.
DR Pfam; PF00903; Glyoxalase; 2.
SO SEQUENCE 275 AA; 28481 MW; DE13C0E714D4397 CRC64;

```

```

alignment_scores:
Quality: 1465.00      Length: 275
Ratio: 5.327          Gaps: 0
Percent Similarity: 100.000      Percent Identity: 99.636

```

```

alignment_block:
US-09-673-254-1/rev x Q9ZAU2

```

```

Align seg 1/1 to: Q9ZAU2 from: 1 to: 275

```

```

2175 GTGACGAGGTTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2126
1 MetThrArgPheAlaProGlyAlaProAlaThrPheAspLeuGlySer 17
2125 CGATGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2076
17 ospValAlaAlaSerAlaAspPheTyrThrGlyLeuPheGlyTyrPhe 34
2075 CCACCGTGTGACGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2026
34 lathValValSerAspProGlyAlaGlyGlyTyrThrThrPheSerSer 50
2025 GACGGGAAGCTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1976
51 AspGlyLysLeuValAlaValAlaValAlaArgHisGlnIleAspThrPro 67
1975 CCACCGGCGCGTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1926
67 HisArgProTyrGlyProGlyAsnAspGlnHisGlyMetProAlaIleT 84
1925 GACCGGTACTTCCGACCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1876
84 rPheValTyrPheAlaThrAspAspAlaAspAlaLeuThrLysArgVal 100
1875 GAGACGGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1826
101 GluThrAlaGlyGlyGluValIleMetThrPrometaAspValLeuGlyLe 117
1825 CGCGCGGATGGCGGTCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGGT 1776
117 uGlyArgMetaValPheAlaAspProAlaGlyAlaAlaPheAlaVal 134

```

```

1775 GCGCGAAGGAGTCAATGAGGCGGAGGTGACGGCGCTGCCGCTCG 1726
134 rPARGysglYalMetgluglYalagluValthrglYalProglYser 150
1725 GTGCGCTGGGTGAGCTGTGACCGGACGATCGGGGCGCCCGGAGCTT 1676
151 ValGlyTTPValIGluLeuValthrglYalIleGlyAlaAlaArgspRh 167
1675 CTACCGGCGGACCTCGGCTGCTCCGGCGGACCGGACGCTGAAGGGCG 1626
167 eTyrTProAlaThrLeuGlYLeuAlaProAlaAspThrGlyLeuLysglY 184
1625 TCACCGACCCGGTGTGACATCGGTGACACACCGGCTCCGACCCAG 1576
84 aLThrsProValTrpHisIleGlyAspThrProValAlaGlyThrGln 200
1575 GAGCTGGGCGGTACCGGCGGTACCGGCGACCTGGGCGCTGTCTGCG 1526
201 GluLeuGlYValThrglYalValAlaArgProHisTTPAlaValLeuPheAl 217
1525 CGTGCACGATCGGACGCGGACGCTCCGGGCGCGCTGTGAATCGGCGCT 1476
217 aValHisAspCysAspAlaThrValAlaArgAlaValIGluLeuGlY 234
1475 CCGTCGAGAACGAGCCCGCGGACACGCGGCGGCGGCGGCGGACCTG 1426
234 eValIGluAsnGluProAlaAspThrProArgGlyAlaArgAlaAspLeu 250
1425 CTGCAACCGGACGCGGCGGCTTCTGCTGTGTAACCTCGGAGGGGTA 1376
251 LeuAspProHisGlyAlaGlyPheSerValValIGluLeuArgGluGly 267
1375 CCGCGGCGGCGGCGGCGGCTGCTCG 1351
267 rProAlaAlaAlaGlyGlyAlaSer 275

seq_name: sp_bacteria:Q55078

seq_documentation_block:
ID Q55078 PRELIMINARY; PRT: 275 AA.
AC Q55078;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
ORA PROTEIN.
ORA.
Streptomyces sp.
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
NCBI_TaxID=1931;
OX 11
RN SEQUENCE FROM N.A.
RC STRAIN=CS;
RX MEDLINE=96236067; PubMed=8655530;
RA Dickens M.L., Strohl W.R.;
RT "Isolation and characterization of a gene from Streptomyces sp. strain
RT C5 that confers the ability to convert daunomycin to doxorubicin on
RT Streptomyces lividans TK24";
RL J. Bacteriol. 178:3389-3395(1996).
DR EMBL: U50973; AAB08048.1;
DR InterPro: IPR004360; Gly_Dleo_dlox.
DR Pfam: PF00903; Glyoxalase; 2.
SQ SEQUENCE 275 AA; 28840 MW; BCDDBBEA60BDE4AF CRC64;

```

```

alignment_scores:
  Quality: 1385.00      Length: 275
  Ratio: 5.149          Gaps: 0
  Percent Similarity: 97.818      Percent Identity: 94.182

alignment_block:
US-09-673-254-1/rev x Q55078

```

```

Align seg 1/1 to: Q55078 from: 1 to: 275
2175 GTGACACGAGTTGGCGCGCGCGCGCGCGCGATGTTGACACTGGTTCCGC 2126
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
2125 CGATGTGCGCGCGCTGCGGCGGCTTGTACACGCGCGGCTGTGCGGTGACCG 2076
17 oAspValAlaAlaSerAlaAspPheTyrThrGlyLeuPheGlyTTPRhA 34
2075 CCACCGTGTGACGACCGCGGCGCGGCGGATATACAGAGTTACACTTC 2026
34 lathValAlaSerAspProGlyAlaGlyTyrThrThrPheSerSer 50
2025 GACGGGAACCTGTGCGCGCGGCTGCGCGCGCGCGCGACAGATGACACCCCTA 1976
51 AspGlyLysLeuValAlaAlaValAlaArgHisGlnIleAspThrProTy 67
1975 CCACCGCGCGGTACGGGCGCGGGAACGACGACGACGCGCATGCGCGCATCT 1926
67 rHisATgProTyTgGlyProGlyAsnAspGlnHisGlyMetProAlaIleT 84
1925 GACCGCTGTACTTGTGCGGACGACGACGCGGCGGCGGCGGCGGCTC 1876
84 rPThrValTyrPheAlaThrAsnAspAlaAspAlaLeuThrLysArgVal 100
1875 GAGACGCGCGCGCGCGGCGGAGTCAATGACTCGGATGAGAGTCTCGGCGCT 1826
101 GluThrAlaGlyGlyAspValIleMetThrProMetAspValLeuGly 117
1825 CGGCGGATGCGGCGGCTTGTGCGGACCGCGGCGGCGGCGGCTTGTGCGGCT 1776
117 uGlyAlaGlyMetAlaValAlaPheAlaAspProSerGlyAlaAlaPheAlaValT 134
134 rPARGysglYalMetgluglYalagluValthrglYalProglYser 150
1775 GCGCGAAGGAGTCAATGAGGCGGAGGTGACGGCGGCTGCCGCTCG 1726
134 rPARGysglYalMetgluglYalagluValthrglYalProglYser 150
1725 GTGCGCTGGGTGAGCTGTGACCGGACGATCGGGGCGCCCGGAGCTT 1676
151 ValGlyTTPValIGluLeuValthrglYalIleGlyAlaAlaArgglYrh 167
1675 CTACCGGCGGACCTGCGGCTGCGGCTGCGGCGGACGCGGAGTGAAGGGCG 1626
167 eTyrTArgAlaThrLeuGlyLeuAlaProAlaAspThrGlyArgLysglY 184
1625 TCACCGACCCGGTGTGACATCGGTGACACACCGGCTCCGACCCAG 1576
184 aLThrsProValTrpHisIleGlyAspThrProValAlaGlyThrArg 200
1575 GAGCTGGGCGGTACCGGCGGTGACCGGCGGACCTGGGCGCTGTCTGCG 1526
201 GluLeuGlYThrThrglYalValAlaArgProHisTTPAlaValLeuPheAl 217
1525 CGTGCACGATCGGACGCGGACGCTCCGGGCGCGCTGTGAATCGGCGCT 1476
217 rValHisAspCysAspAlaThrValAlaArgAlaValIGluLeuGlY 234
1475 CCGTCGAGAACGAGCCCGCGGACACGCGGCGGCGGCGGCGGACCTG 1426
234 eValIGluAsnGluProValAspThrProArgGlyAlaArgAlaAspLeu 250
1425 CTGCAACCGGACGCGGCGGCTTCTGCTGTGTAACCTCGGAGGGGTA 1376
251 LeuAspProHisGlyAlaGlyPheSerValValIGluLeuArgGluAlaTy 267
1375 CCGCGGCGGCGGCGGCGGCTGCTCG 1351
267 rProAlaAlaAlaAspGlyAlaSer 275

seq_name: sp_bacteria:Q92A01

```

seq_documentation_block:
 ID Q9ZAU1 PRELIMINARY; PRT; 287 AA.
 AC Q9ZAU1;
 DT 01-MAY-1999 (TREMBlrel. 10, Created)
 DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE DAUNORUBICIN C-13 KETOREDUCTASE.
 GN DNRU.
 OS Streptomyces peuceletius.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=1950;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 29050;
 RX MEDLINE=99084971; PubMed=9864344;
 RA Lemovskaya N., Otten S.L., Doi-Katayama Y., Foonstein L., Liu X.-C.,
 RA Takatsu T., Invenit A., Filippi S., Torti F., Colombo A.L.,
 RA Hutchinson C.R.;
 RA "Doxorubicin overproduction in Streptomyces peuceletius: cloning and
 characterization of the dnrV ketoreductase and dnrV genes and the doxV
 cytochrome P-450 hydroxylase gene."
 J. Bacteriol. 181:305-318(1999).
 -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
 (SDR) FAMILY.
 CC EMBL: U77891; AAD04717.1; -;
 DR InterPro: IPR002198; ADH_short.
 DR Pfam: PF00106; adh_short; 1.
 KM Oxidoreductase.
 SQ SEQUENCE 287 AA; 30632 MW; 4F9448EAE55F6191 CRC64;

alignment_scores:
 Quality: 1014.00 Length: 196
 Ratio: 5.173 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-673-254-1/rev x Q9ZAU1 ..

Align seg 1/1 to: Q9ZAU1 from: 1 to: 287

```

2869 CGGTACCCGCGATGATGTCATGGCCAGCAAGCGCGCGCATGTTCTG 2820
|||||
92 ArgTyrProArgIleAspValMetAlaSerAsnAlaGlyGlyMetPheTr 108
2819 GTCCGGCACACGACCGACGGGTGCGACCGACCATCCAGGCGCAATC 2770
|||||
108 pSerArgThrThrThrGlnAspGlyPheGlnAlaThrIleGlnValAsnH 125
2769 ACCTGCGAGGCTTCTGCTGGCAGCGGTGCTGCGGAGCGGCTCGCGGAC 2720
|||||
25 IsteuAlaGlyPheLeuLeuAlaArgLeuAlaArgGlnArgLeuAlaGly 141
19 GGGCGGTGATCTCACTCTGTCGACGCGTACACCGACGGCGGATCGA 2670
|||||
142 GlnArgLeuIleLeuThrSerSerAspAlaTyrThrGlnGlyArgIleAs 158
2669 CCGGAGACGACCTCAACGGCGGACCGCTACCGGTACGCGCGCGCGGCGT 2620
|||||
158 pProAspAspLeuAsnGlyAspArgHisArgTyrSerAlaGlyGlnAlaTr 175
2619 ACGGACGCTCAAAACAGGCGCAACATCATGACCGCGGCGAGCGCCGACG 2570
|||||
175 YrGlyThrSerTyrGlnAlaAsnIleMetThrAlaAlaGlnAlaIleArg 191
2569 GCGTGGCCGCGACGTGCTGGCGGTACGCTATACACCGGTGAGGTCCGAC 2520
|||||
192 ArgTyrProAspValLeuAlaValSerTyrHisProGlyGlnValArgTh 208
2519 CCGCATCGAGAGGGGACGCTGCGCTGCTACTCTCCGTTCAACCCCT 2470
|||||
208 rArgIleGlyArgGlyThrValAlaSerSerTyrPheArgPheAsnProp 225

```

```

2469 TCCTGGCGTCCGGCGGAGGCGCGACACCTGCTGCTGGCGTCC 2420
|||||
225 heuLeuArgSerAlaAlaGlyGlyAlaAspThrLeuValTrpLeuAlaSer 241
2419 GCGCGCGCGGAGAGTTGACACGAGCGGCGCTACTACAGCAGCGCGGCGCT 2370
|||||
242 AlaProAlaGlnGlnLeuThrThrGlyGlyTyrTyrSerAspArgIle 258
2369 GTCCCGGTGAGCGGCGCGACCGCGCGGCGCTGCGGCGAAGCTCT 2320
|||||
258 uSerProValSerGlyProThrAlaAspAlaGlyLeuAlaAlaTyrLeuTr 275
2319 GGGAGCGCGCGCGCGCGCGTGGCGGACACGCGGCGAC 2282
|||||
275 rPoluAlaGlyAlaAlaAlaValAlaGlyAspThrAlaHis 287

```

seq_name: sp_bacteria:Q53877

seq_documentation_block:
 ID Q53877 PRELIMINARY; PRT; 287 AA.
 AC Q53877;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE KETOREDUCTASE.
 OS Streptomyces sp. C5.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=45212;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-C5;
 RX MEDLINE=96236066; PubMed=8655529;
 RA Dickens M.L., Ye J., Strohl W.R.;
 RT "Cloning, sequencing, and analysis of aklaviketone reductase from
 Streptomyces sp. strain C5."
 RL J. Bacteriol. 178:3384-3388(1996).
 -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
 (SDR) FAMILY.
 CC EMBL: U43704; AAB08016.1; -;
 DR InterPro: IPR002198; ADH_short.
 DR Pfam: PF00106; adh_short; 1.
 KM Oxidoreductase.
 SQ SEQUENCE 287 AA; 30799 MW; 8FEB9D1BAFE1E78 CRC64;

alignment_scores:
 Quality: 980.00 Length: 195
 Ratio: 5.026 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 96.410

alignment_block:

US-09-673-254-1/rev x Q53877 ..

Align seg 1/1 to: Q53877 from: 1 to: 287

```

2869 CGGTACCCGCGATGATGTCATGGCCAGCAAGCGCGCGCATGTTCTG 2820
|||||
92 ArgTyrProArgIleAspValMetAlaGlyAsnAlaGlyGlyMetPheTr 108
2819 GTCCGGCACACGACCGGAGGCGGTGAGGCGCACATCCAGGCGCATC 2770
|||||
108 pSerArgThrThrThrGlnAspGlyPheGlnAlaThrIleGlnValAsnH 125
2769 ACCTGCGAGGCTTCTGCTGGCAGCGGTGCTGCGGAGCGGCTCGCGGAC 2720
|||||
125 IsteuAlaGlyPheLeuLeuAlaArgLeuAlaArgGlnArgLeuAlaGly 141
2719 GGGCGGTGATCTCACTCTGTCGACGCGTGTGCGGAGCGGCTCGCGGAC 2670
|||||
142 GlnArgLeuIleLeuThrSerSerAspAlaTyrThrGlnGlyArgIleAs 158

```

```
2669 CCGGAGACCTGACGGCGACCGCTACCGCTACAGCCGCGGCGAGCGT 2620
|||||
158 pProAspAspLeuAsnGlyAspArgHisArgTyrSerAlaGlyGlnAlaT 175
|||||
2619 ACGGACAGTCCAAACAGCCCAACATCATGACCGCGGAGCGCGCCAGG 2570
|||||
175 yrlGlyThrSerLysGlnAlaAsnIleMetThrAlaThrGlnAlaAlaArg 191
|||||
2569 CGCTGGCCGGACGCTGCGGCGGTACGCTATCAACCCGCGTGGAGTCCGAC 2520
|||||
192 ArgTrpProAspValLeuThrValSerTyrHisProGlyGlnValArgTh 208
|||||
2519 CCGCATCGGACGGGGGACGCGTCCCTCTACTTCCGGTTCAACCCCT 2470
|||||
208 rArgIleGlyArgGlyThrValAlaSerThrTyrPheArgPheAsnProp 225
|||||
69 TCCTGGGCTCCGCGGGAAGGGCGCGACACCCCTCGTGGCTGGCGTCC 2420
|||||
225 heLeuArgSerAlaAlaLysGlyAlaAspThrLeuValTrpLeuAlaAla 241
|||||
2419 GCGCCGCGCGAGAGTTGACACGAGGCGGCTACTACAGCGACGCGGCGT 2370
|||||
242 AlaProAlaGlnGluLeuThrThrGlyGlyTyrTyrSerAspArgArgLe 258
|||||
2369 GTCCCGGTGAGCGCGCGACCGCGACCGCGGCTCGCGGCGAAGCTCT 2320
|||||
258 uSerProValSerGlyProThrAlaAspAlaGlyLeuAlaAlaLysLeuT 275
|||||
2319 GGGAGGCGCGCGCGCGCTCGCGGACACCGCG 2285
|||||
275 rPeLnuAlaSerAlaAlaAlaValaGlyAspThrAla 286
```

